Supplementary Material

High-resolution interrogation of functional elements in the noncoding genome

Neville E. Sanjana[1,2,*,†,‡], Jason Wright[1,2,*], Kaijie Zheng[1,2], Ophir Shalem[1,2], Pierre Fontanillas[1], Julia Joung[1,2], Christine Cheng[1,3], Aviv Regev[1,3], Feng Zhang[1, 2,†]

[1] Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA

[2] McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[3] Howard Hughes Medical Institute, David H. Koch Institute of Integrative Cancer Biology, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

* These authors contributed equally to this work.

† Correspondence should be addressed to: nsanjana@nygenome.org (N.S.) and zhang@broadinstitute.org (F.Z.).

‡ Present address: New York Genome Center, New York, NY 10013, Department of Biology, New York University, New York, NY 10012.

MATERIALS AND METHODS

Noncoding library design and cloning

To design the noncoding libraries for *NF1*, *NF2*, and *CUL3*, we selected regions of 100 kb flanking the coding sequence for both of the most highly expressed RefSeq isoforms as determined by RNA-seq quantification in BRAF-mutant A375 melanoma cells (*NF1* primary: NM_001042492, *NF1* alternate: NM_000267; *NF2* primary: NM_000268, *NF2* alternate: NM_016418; *CUL3* primary: NM_003590, *CUL3* alternate: NM_001257197). We also included the 5′ and 3′ untranslated regions (UTRs). For these regions, we identified all Cas9-targetable sites on both strands, i.e. those containing the protospacer-adjacent motif (PAM) `NGG`. We eliminated sgRNAs with potential off-targets elsewhere in the genome as described previously (*26*), which yielded 18,315 sgRNAs with the following median distances between neighboring sgRNAs for each library: *NF1* 17 bp, *NF2* 12 bp, *CUL3* 19 bp. Genomic sequences were retrieved using the UCSC Genome Browser (*hg19*) and Galaxy. Custom Python scripts for sgRNA guide design are available on GitHub (http://github.com/nsanjana/BashRegion).

The sgRNA sequences were synthesized as single-stranded oligonucleotides on a CustomArray synthesizer, PCR amplified using Phusion Flash (ThermoFisher Scientific F548L) polymerase (15 cycles), and Gibson cloned into a guide-only lentiviral vector (lentiGuide-Puro, Addgene 52963).

Vemurafenib pooled lentiviral production and screening

The vemurafenib resistance screen was conducted similarly to a previously described genome-wide CRISPR screen (*16*). Lentivirus was produced via transfection of library plasmid with appropriate packaging plasmids (psPAX2: Addgene 12260; pMD2.G: Addgene 12259) using Lipofectamine 2000 and Plus reagent (ThermoFisher Scientific, 11668019 and11514015) in HEK293FT (ThermoFisher Scientific, R70007). At 3 days post-transfection, virus was collected and passed through a 0.45 um filter and stored at -80C until use (supernatant, unpurified virus).

For the screen, A375 human melanoma cells (ATCC CRL-1619) were cultured in RPMI-1640 media (ThermoFisher Scientific 61870127) with 10% fetal bovine serum (Seradigm 1500-500) and no antibiotics (“R10 media”). To first introduce Cas9, A375 was transduced with a Cas9-

expressing lentivirus (lentiCas9-Blast, Addgene 52962) and selected for 7 days with 10 ug/mL blasticidin. Resistant cells were expanded and transduced with the *CUL3* library (lentiGuide-Puro) pooled lentivirus in 2 separate infection replicates with $3.45x10^7$ cells per infection replicate using a standard spinfection protocol. After 24 hours, cells were selected with 1 ug/mL puromycin for 7 days, resulting in ~30% cell survival. The overall representation was ~1000 cells per construct (830 in replicate 1 and 1130 in replicate 2) with ~83% of surviving cells receiving a single sgRNA construct (see (*9*) for details of Poisson infection model and single-infection percentage calculation).

After 7 days, we removed puromycin and split cells into separate flasks with either 2uM vemurafenib (PLX4032, Selleckchem S1267 in DMSO) or an equal volume of DMSO. At this point, a representative sample of $3x10^7$ cells from each infection replicate was frozen at -20C as an early time point ("Day 0") for screen readout. All flasks were either passaged or had fresh media added every 2 days. At day 14 after addition of vemurafenib/DMSO, the screen was terminated and $1\text{-}3x10^7$ cells were frozen at -20C for each condition/replicate ("Day 14").

Screen readout and data analysis

For each timepoint/sample, genomic DNA was extracted following a modified salting-out precipitation method described previously in detail (*9*). The sgRNA readout was performed using two rounds of PCR (*16*). For the first PCR step, a region containing the sgRNA cassette in the lentiviral genomic integrant was amplified from extracted genomic DNA using the PCR1 primers in table S2.

For each sample, we performed 12 duplicate PCR reactions with 3 ug of gDNA in each reaction (total gDNA = 36 ug per sample for representation of $\sim 5x10^6$ cells). We pooled the unpurified PCR products and used the mixture for a single second PCR reaction per biological sample. This second PCR adds on Illumina sequencing adaptors, barcodes and stagger sequences to prevent monotemplate sequencing issues. Complete sequences of the 12 forward and 12 reverse Illumina PCR2 readout primers used are shown in table S2.

PCR reactions were performed using Phusion Flash (ThermoFisher Scientific F548L) polymerase following the default protocol with an annealing temperature of 62C and 20 cycles.

Amplicons from the second PCR were pooled in equimolar ratios (by gel quantification) and then purified using a QiaQuick PCR Purification kit (Qiagen 28104). Purified products were loaded onto a 2% E-gel EX and gel extracted using a QiaQuick Gel Extraction kit (Qiagen 28704). The concentration of the gel-extracted PCR product was gel quantified using the Low-Range Quantitative Ladder (ThermoFisher Scientific 12373031) and then diluted and sequenced on an Illumina MiSeq using a v3 kit (Illumina MS-102-3001).

Reads were demultiplexed using FASTX-Toolkit and aligned to the designed sgRNAs using `bowtie` (with parameters `-v 1 -m 1 --norc`) (*27*). Read counts were imported into R/RStudio and normalized within each sample. Over all 3 screens, the correlation between the infection replicates was relatively high: $r = 0.93 \pm 0.02$ (mean ± s.d.) for DMSO Day 14 and $r = 0.80 \pm 0.02$ for Vemu Day 14. All plots and analyses are from the average of the two infection replicates, unless indicated otherwise.

Estimation of false positive/negative rates and optimal enrichment cutoff

We quantified the false positive rate using two independent metrics and found similar results:

1) quantifying the enrichment of coding regions of genes not implicated in vemurafenib resistance in the genome-scale CRISPR screen (*16*),
2) analyzing the enrichment of the non-targeting control sgRNAs embedded in each library.

Enrichment of negative control gene coding regions: We examined CDS-targeting sgRNAs from 11 neighboring genes which overlap with our screen regions. Using our A375 RNA-seq dataset, we found that these genes were, as a group, expressed at similar levels to the genes previously implicated in vemurafenib resistance (fig. S2A). For all of these putative negative controls, the average enrichment for each gene was always less than zero, providing further evidence that the false positive rate is low (fig. S2B). In our validation (Figure 3), we selected putative functional elements by looking for multiple enriched sgRNAs that were <500 bp apart. Of the 782 sgRNAs targeting the CDS of these genes, only 11 sgRNAs (1.4%) were not depleted (fig. S2C). Of these enriched sgRNAs, only 4 of them were within 500 bp of another enriched sgRNA, yielding a false positive rate of 4 out of 782 (<1%).

Enrichment of non-targeting control sgRNAs: All three libraries include a set of non-targeting sgRNAs, which have guide sequences designed to **not** target the human genome. These are similar to sgRNAs we have used as non-targeting negative controls in genome-wide libraries (*9*, *26*). Since these sgRNAs should have no effect on drug resistance, any enrichment is assumed to be a false positive. Using the same criteria to identify enriched sgRNAs as shown in Figure 1B-D (>4 standard deviations from control), we find a false positive rate of 5% or less (fig. S2D).

Receiver-operator characteristic (ROC) analysis to determine enrichment cutoff: Both estimators of the false positive rate (enrichment of coding regions of nearby but uninvolved in vemurafenib resistance genes and enrichment of sgRNAs that do not target the human genome) indicate a low false positive rate of 0.5% to 5% depending on the estimator and screen dataset (fig. S2E). We varied the cutoff for enrichment (in units of standard deviations from the sgRNA distribution in the control conditions) to select an optimal value.

To do this, we computed the true positive rate as the fraction of enriched CDS-targeting sgRNAs (for *NF1*, *NF2* and *CUL3*) (fig. S3A, *left*) and the false positive rate as the fraction of enriched non-targeting control sgRNAs (fig. S2D). For the true positive rate (=1 - the false negative rate), it would be ideal to use sgRNAs that target noncoding regions but, since we do not know which noncoding regions are *a priori* involved in gene regulation, we have used CDS-targeting sgRNAs as a surrogate. It is possible that the rules for functional disruption of a noncoding region differ from those for coding regions. Based on the inflection point of the average ROC curve across the 3 screens (fig. S2F), we chose an enrichment cutoff of 4 standard deviations from the sgRNA distribution in the control condition.

RNA-sequencing (RNA-seq) from human A375 (V600E BRAF) melanoma cells

RNA from A375 cells was harvested using the RNeasy Plus Mini Kit (Qiagen 74134) and prepared with TruSeq Stranded Total RNA Kit with Ribo-Zero Gold (Illumina RS-122-2303). Samples were deep-sequenced on the Illumina NextSeq platform (>20 million reads per condition). A Bowtie index was created based on the human *hg19* UCSC reference genome and

RefSeq known transcriptome, and RSEM v1.27 was run with default parameters to align paired-end reads to this index to estimate expression levels.

Chromatin conformation capture (3C) with droplet digital PCR (ddPCR) quantification

To map physical interactions between distal sites and the *CUL3* promoter in A375 cells, we made three independent 3C libraries using different 6-cutter restriction enzymes (*EcoRI*, *BglII*, and *HindIII*). For each library, $1x10^8$ log-phase A375 cells were cross-linked, digested and ligated using a standard protocol from Job Dekker and colleagues (*28*). For quantitative PCR of the purified genomic DNA from the 3C libraries, we designed unidirectional primers flanking each cut site in the region using Rebase (New England Biolabs) (see table S3 for primer sequences and enzyme cut sites).

As 3C results are influenced heavily by differences in primer amplification efficiency, we used droplet digital PCR (ddPCR) with EvaGreen to quantify interaction frequencies. For each droplet (~20,000 per PCR reaction), a digital readout of amplification/no-amplification is used after saturation PCR (40 cycles). For each library, we optimized over a range of input template concentrations to find the ideal template concentration for droplet quantification (i.e. sufficient positive and negative droplets for Poisson estimation). ddPCR reactions were performed in triplicate and we found good agreement between the three independent libraries. Overall enrichment was plotted by smoothing the combined data from the three independent 3C libraries with a Gaussian kernel with a standard deviation equal to half of the average distance between restriction enzyme cut sites (average distance = 4.3 kb, σ = 2.15 kb). For the 12 strongest interactions, we separately PCR amplified and Sanger sequenced the products to validate that they contained the predicted junction.

To correlate enrichment with 3C interaction frequency, we first created windows across the library region because the resolution of 3C is much coarser than the resolution of the sgRNA library. We set the size of each window equal to the average distance between 3C restriction enzyme cut sites (4.3 kb) with a ~75% overlap between windows (i.e. one window every kilobase). For each window, we calculated the average enrichment ($\log_2$ Vem/DMSO) of the sgRNAs in the window and used this quantity as the enrichment score of the window. Typically, each 4.3 kb window contained ~100 sgRNAs. For each 3C interaction, we identified the closest

4.3 kb window (defined as the window center) and assigned its enrichment score to the 3C interaction. In our analysis, we looked at all enrichment scores with a particular *minimum* normalized 3C interaction and did this across the entire range of normalized 3C interactions.

Assay for transposable and accessible chromatin sequencing (ATAC-seq)

For ATAC-seq, human melanoma A375 (ATCC CRL-1619), mammary gland adenocarcinoma MCF-7 (ATCC HTB-22), and glioblastoma U87-MG (ATCC HTB-14) cells were cultured in R10 media (RPMI-1640 + 10% FBS, as described above). For each line, $5x10^4$ cells in log-phase growth were harvested using an existing ATAC library preparation protocol with minor modifications (*21*). Library quality was validated using an Agilent TapeStation before pooling barcoded samples and sequencing using an Illumina NextSeq with 36 bp paired-end reads. Each sample was sequenced to a depth of ~75M reads.

Samples were aligned using `bowtie` (with parameters `--chunkmbs 256 -p 24 -S -m 1 -X 2000`) to the human genome reference sequence (*hg19/GRCh37*). The resulting BAM files were subset using `samtools` to the region our sgRNA library targets (*hg19* coordinates: chr2: 225,234,905-225,550,015). For quality control, we measured the duplicate read rate using Picard-Tools MarkDuplicates (10-30%) and also the mitochondrial read rate (<5%) (*29*). Aligned BAM files were converted to BEDgraph format using `bedtools` (*30*) and imported for analysis into R/RStudio.

DNAse I hypersensitivity and chromatin immunoprecipitation sequencing (ChIP-seq) datasets

For comparison with screen enrichment, we used DNAse I hypersensitivity and ChIP-seq data from the ENCODE project (*2*). DNAse I hypersensitivity data for Colo829 melanoma, MCF7 mammary gland, and Gliobla D54 glioblastoma data is from the OpenChrom/Duke University collection. All ChIP-seq data is from K562 cells: YY1 and ZNF263 are from the Stanford/Yale/USC/Harvard dataset; CTCF is from the Open Chrom/UT Austin dataset; and c-Fos and JunD are from the U. Chicago dataset. All files were downloaded as variable-step wig format using the UCSC Table Browser.

Fold enrichment of sgRNAs near chromatin accessibility and sequence conservation peaks

PhastCons data for primates (n=10 animals), placental mammals (n=33), and vertebrates (n=46) were downloaded from UCSC for *hg19*. To calculate the fold enrichment of the sgRNAs in proximity to other molecular hallmarks (DNAse-seq, ATAC-seq, conservation), we examined the average sgRNA enrichment of sgRNAs near the peaks of these molecular hallmarks. We then followed a Monte Carlo procedure: We randomized the peak locations over the screen region and recomputed the average sgRNA enrichment. We performed 10,000 random reshufflings of the peak locations over the screen region to get a distribution of average sgRNA enrichments. Fold enrichment is the ratio of the average sgRNA enrichment using the actual peak locations divided by the mean of the Monte Carlo distribution (average sgRNA enrichment with reshuffled peak locations). Significant sgRNA enrichment at actual peak locations was $p<0.05$ in the Monte Carlo distribution.

Array validation of primary screen hits

For individual (array) validation of noncoding sgRNAs, we first identified enriched sgRNAs as given by the normalized $\log_2$(Vemu/DMSO) read ratio (approx the top ~5% of the library). In order to have high confidence in these sgRNAs, we used the *minimum* of the two infection replicates for the normalized $\log_2$(Vemu/DMSO) read ratio. From this group, we eliminated any sgRNAs that did not have another similarly enriched sgRNA within 500 bp. This ensures that putative noncoding functional elements were supported by the presence of at least 2 enriched sgRNAs. Of the enriched sgRNAs in the *CUL3* screen, 78% (179 out of 230 sgRNAs) had another enriched sgRNA within 500 bp. From this group, we picked 25 sgRNAs distributed across different genomic regions for individual validation (see table S4 for a list of sgRNA sequences). We also included 3 exon-targeting and 3 non-targeting sgRNAs to serve as positive and negative controls, respectively.

For each sgRNA, standard desalted short oligonucleotides (Integrated DNA Technologies) were annealed, phosphorylated and cloned into a lentiviral vector (lentiCRISPRv2, Addgene 52961) that contained Cas9 and an sgRNA cassette. For each sgRNA, A375 cells were transduced with

lentiviral supernatants. After 24 hours, media was replaced with R10 with 1 ug/mL of puromycin. Viral volumes were titered such that 20-40% of cells survived after puromycin selection. After selection and expansion for 7 days in puromycin, cell were plated for DNA/RNA extraction, vemurafenib resistance, or ChIP assays.

Deep sequencing after CRISPR mutagenesis

After 7 days of puromycin selection, A375 cells transduced with individual lentiCRISPRv2 sgRNAs were plated in 2 replicate wells ($2x10^3$ cells/well) in 96-well plates. Cells were plated in either R10+DMSO or R10+vemurafenib (2 uM). After 4 days (70-90% confluent), we extracted gDNA from all wells (Illumina/Epicentre QuickExtract QE09050) and performed amplification and deep sequencing as previously described (*16*). Briefly, for each sgRNA target site, we designed PCR primers to amplify genomic regions surrounding the site (100-200 bp amplicons) and to add universal handles for the second stage of amplification (see table S5 for all deep sequencing primers). We then used a second PCR step to add sequences needed for Illumina sequencing and sample barcoding. We pooled all samples together and sequenced them on a MiSeq using a 250bp single-end read (Illumina MS-102-2002).

To measure indel length and eliminate any potential off-target or primer-dimer reads, we first identified our genomic (first PCR step) primers in each read. We then checked that each read contained at least 5 bases beyond the 3′end of each of the genomic primers. Typically, 80-90% of demultiplexed reads matched this criterion. Reads matching this criterion were used to measure indel length by comparing distances between the identified primer-adjacent sequences with those in the reference sequence. Multiple alignment analysis for specific sgRNAs was done using Geneious's iterative *k*-mer multiple alignment tool (Geneious 6.1.7).

RNA extraction and ddPCR quantification of *CUL3* expression

After 7 days of puromycin selection, A375 cells transduced with individual lentiCRISPRv2 sgRNAs were plated in 3 replicate wells ($2x10^3$ cells/well) in 96-well plates. After 4 days (70-90% confluent), RNA was extracted using a homemade version of a rapid lysis kit for quantitative PCR, as described previously (*31*). Briefly, for each condition cells were lysed by

proteinase-K lysis solution and DNase I treatment in appropriate lysis buffers in 96-well plates. For each sample, cDNA was generated with reverse transcriptase using random hexamers and oligo-dT.

To measure *CUL3* expression, we used a ddPCR-based TaqMan assay (dual-label probe hydrolysis by *Taq* polymerase exonuclease activity). We first tested two different *CUL3* TaqMan probe designs to determine which one provided better separation between amplification/no-amplification droplets. Of the two probes tested (*Hs00180183_m1* and *Hs00950986_m1*), we found that *Hs00950986_m1* achieved the best separation in the droplet analysis and used it for all *CUL3* expression assays as the FAM channel probe (ThermoFisher Scientific). For normalization, we used a TaqMan probe for *TBP* (*TATA*-box binding protein) in the VIC channel (ThermoFisher Scientific 4326322E). In each 24 ul reaction, we used 9.6 ul of the cDNA produced by our homemade RNA extraction/reverse transcription protocol and 1.2 ul of each probe (*CUL3* and *TBP*). Droplets were formed using the 96-well droplet generator (BioRad AutoDG), thermocycled following BioRad's standard TaqMan protocol, and then analyzed using a two-channel ddPCR reader (BioRad QX200). *CUL3* expression was first normalized by *TBP* expression in each well and then normalized across samples using the expression level from the average of 3 different non-targeting sgRNAs.

Vemurafenib resistance assay

After 7 days of puromycin selection, A375 cells transduced with individual lentiCRISPRv2 sgRNAs were plated in 8 replicate wells ($2x10^3$ cells/well) in 96-well black-bottom plates. After 24 hours, the media was replaced with R10 with 2uM vemurafenib (4 wells) or R10 with an equal volume of DMSO (4 wells). Drug/vehicle media was replaced every other day. After 3 days, cell viability was measured using CellTiter Glo (Promega) using the following protocol: After cells were equilibrated to room temperature (30 minutes), media was aspirated and replaced with CellTiterGlo reagent diluted 1:4 in PBS. Cells were placed on an orbital shaker for 2 minutes and then incubated for an additional 10 minutes before luminescence measurement (1s integration time) on a plate reader (Biotek Synergy H1). Linear fit and correlation with *CUL3* expression ($r_{noncoding}$ = -0.54, p = 0.005, t-test for Pearson correlation) only includes noncoding sgRNAs and does not include coding-targeting or non-targeting sgRNAs. With coding-targeting

sgRNAs and non-targeting sgRNAs included, the correlation coefficient is r_{all} = - 0.65 ($p < 10^{-4}$).

Lentiviral CUL3 overexpression

To create the CUL3 lentiviral transfer plasmid, we PCR amplified CUL3 from an A375 cDNA library using primers specific to a particular isoform (isoform 1, RefSeq accession NM_003590) and performed a Gibson ligation into lentiCas9-Blast (Addgene 52962) cut with *AfeI* and *BamHI*. A similar procedure was used to create an EGFP control plasmid. After 7 days of puromycin selection, A375 cells transduced with individual lentiCRISPRv2 sgRNAs were transduced with either CUL3 or EGFP blasticidin-selectable lentiviruses (pLenti-EFS-CUL3-P2A-Bla or pLenti-EFS-EGFP-P2A-Bla). Since the Cas9 sgRNAs only target noncoding regions, expression of CUL3 from the transgene should not be affected since it has no introns (ORF clone) and is driven by a different promoter (EF1alpha short/EFS). We then selected for cells transduced with either CUL3 or EGFP using R10 supplemented with 10 ug/ml of blasticidin for 7 days before plating for the vemurafenib resistance assay using CellTiter Glo (see above).

Chromatin immunoprecipitation (ChIP) for histone modifications and transcription factors

After 7 days of puromycin selection, A375 cells transduced with individual lentiCRISPRv2 sgRNAs were plated in T-225 flasks and grown to 70-90% confluence (6 days). At this point, chromatin fixation was initiated by adding formaldehyde directly to the growth media (final concentration 1%) and incubating at 37 C for 10 minutes. The entire two-day ChIP procedure was performed using the Magna ChIP HiSens Chromatin Immunoprecipitation Kit (Millipore 1710460), as specified in the manufacturer's protocol. Sonication conditions were 2 rounds of 10 minutes of pulse sonication (30s on-off cycles, high frequency) in a rotating water bath sonicator (Diagenode Bioruptor) with 5 minutes on ice between each round.

Using BatchPrimer3, we designed primers centered on the sgRNA target site with a target amplicon size of 80-120 bp (table S6). Antibody concentrations were optimized individually for each antibody (table S7). Droplet digital PCR (ddPCR) reactions using EvaGreen (BioRad 1864034) were used to quantify changes between input, histone/TF ChIP, and IgG ChIP samples for A375 cells transduced with specific sgRNAs and untransduced A375 cells. We first used the

IgG ChIP (negative control) to make sure that there was minimal background. For all histones/TFs, we also designed primers using the same method (BatchPrimer3) for positive control regions (unrelated to the *CUL3* locus) and verified that they were unchanged after editing by validation sgRNAs. We calculated the percent change in ChIP signal after genome editing by normalizing each ChIP sample to its corresponding input sample and then comparing the normalized ChIP between A375s transduced with specific sgRNAs and untransduced (control) A375 cells. Significance testing was performed using Student's *t*-test.

Transcription factor motif prediction

At validation set sgRNA sites, transcription factor binding site prediction was carried out by using 100 bp of genomic sequence centered on each cut site. This sequence was entered into the JASPAR database (http://jaspar.genereg.net), a non-redundant set of transcription factor binding profiles derived from published datasets of transcription factors binding sites (*32*). For programmatic access to the JASPAR database and relative score calculations, we used the R/Bioconductor package `TFBSTools` (*33*). Candidate transcription factor binding sites were identified by overlap of sgRNA cut sites with predicted motifs using a relative profile score threshold of 80% (i.e. the default JASPAR setting). Motif scores are JASPAR relative scores, which is defined as 1 for the maximum-likelihood sequence. The relative score is the sum of the $\log_2$ normalized position-weight matrix probabilities for each base relative to the sum of the $\log_2$ normalized maximum likelihood (i.e. max scoring) sequence for the position-weight matrix (*34*).

SUPPLEMENTARY TEXT

Mutagenesis of distal and proximal histone acetyl-transferase p300 binding sites

Given the observed changes in *CUL3* expression and the surrounding epigenetic environment (Fig. 3), we explored the impact of noncoding mutagenesis on histone-modifying protein occupancy and activity. Two sites targeted by validation sgRNAs occupy local peaks of enrichment for a histone acetyl-transferase and transcriptional co-activator, p300 (fig. S9A). p300 expression and localization is prognostic in *BRAF* mutant melanoma (*35*), and histone deacetylase inhibitors have been shown to work synergistically with vemurafenib to treat cancer (*36*). Although the two p300 sites are separated by ~22 kb, our 3C data indicates a strong interaction (fig. S9B) that could bring the distal p300 site close to the proximal p300 site, which overlaps with the promoter region of *CUL3* (fig. S9C). To explore if sgRNAs targeting these p300 sites alter occupancy and acetylation, we performed ChIP-ddPCR at both sites using antibodies for p300 and H3K27ac. After genome modification with the respective sgRNAs, we found a ~50% loss of p300 occupancy at each site (fig. S9D) and a similar decrease in *CUL3* expression (fig. S9E). In addition, after editing at the distal site, we detect a 93% loss of H3K27ac at that site (fig. S9F) while levels of H3K27ac at a positive control region distant from the *CUL3* locus were unchanged (fig. S8C). Furthermore, we find a 75% decline in H3K27ac at the promoter site after editing at the distal site (fig. S9F). These findings suggest that a distal p300 binding site contributes to maintenance of promoter-proximal histone acetylation, which promotes gene expression.

Extended description of transcription factor binding sites targeted in Figure 4

Identification of other noncoding elements, such as transcription factor binding sites, that regulate *CUL3* may provide new mechanistic insights into resistance or identify therapeutically tractable targets. To identify candidate transcription factors whose binding sites might be disrupted, we further analyzed via next generation sequencing specific sgRNA target sites after editing and queried these target sites for disruption of known transcription factor motifs using the JASPAR database of transcription factors. At four sgRNA target sites, the canonical transcription factor motifs for Yin Yang 1 (YY1), Zinc Finger Protein 263 (ZNF263), CCCTC-binding factor (CTCF) and activation protein 1 (AP-1) complex were severely disrupted after editing (Figs. 4A-

D, fig. S10, S11). Based on these observations we hypothesized that mutations within these binding sites abrogate transcription factor recruitment leading to loss of *CUL3* expression and increased vemurafenib resistance. To test these hypotheses, we compared ChIP-ddPCR enrichment of each transcription factor in cells transduced with a sgRNA from our validation set and in control cells (transduced with a non-targeting sgRNA). In the 5′ UTR, two sgRNAs (5′-UTR sg1, sg2) spaced <50 bp apart overlap a YY1 ChIP-seq peak (Fig. 4A). YY1 is a multifunctional transcription factor capable of both gene activation and repression and its overexpression has been observed in various human malignancies (*37*, *38*). Analysis of the region using the JASPAR motif and scoring algorithm identifies a canonical YY1 motif with 100% relative score (i.e. the unedited reference sequence perfectly matches the maximum likelihood YY1 motif) (fig. S11A) (*32*, *34*). After editing with 5′-UTR sg1, the average relative score for the YY1 motif falls to 82% (n = 1000 sequencing reads), which is nearly the same as the average score for this motif in random DNA sequences (n = 1000 length-matched random sequences) (fig. S11B). Furthermore, we found an increased disruption of the YY1 motif in vemurafenib-treated cells versus vehicle treatment (fig. S12), suggesting that vemurafenib treatment enriches for binding site-damaging mutations. ChIP-ddPCR shows that both sg1 and sg2 decrease YY1 binding, and sg2 (which cuts closer to YY1) more efficiently disrupts YY1 binding than sg1 (67% vs. 26%) (Fig. 4E). In addition, both sg1 and sg2 significantly decrease *CUL3* expression (Fig. 4E). Although there is a large difference in YY1 binding, it is surprising that there is a similar decrease in *CUL3* expression. This could be explained by other altered regulatory elements disrupted by sg1 (i.e. aside from YY1) or that both sgRNAs disrupt enough YY1 binding to saturate the possible YY1-based downregulation of *CUL3*. Similarly, 2 sgRNAs in the first intron of *CUL3* (Intron-sg1, sg2) spaced 30 bp apart overlap a ZNF263 ChIP-seq peak (JASPAR relative score: 89%) (Fig. 4B). Both sg1 and sg2 result in a significant decrease in ZNF263 occupancy via ChIP-ddPCR and a decrease in *CUL3* expression (Fig. 4F).

Although we observe a bias in the presence of regulatory elements 5′ of the transcription start site, we did find several highly enriched sgRNAs downstream of *CUL3,* including two sgRNAs that overlap with AP-1 complex binding sites (distal 3′ sg1, sg2) and another sgRNA that targets a CTCF binding site (CTCF sg1) (Fig. 4C, D). The CTCF sg1 site lies ~30 kb from the 3′ end of *CUL3* and overlaps with non-tissue specific CTCF ChIP-seq peaks of enrichment (Fig. 4C).

CTCF sites are frequently mutated in cancer, and CTCF has been shown to act as an activator, repressor, insulator and mediator of chromatin organization and chromatin loop formation (*39*, *40*). Although we did not find evidence for a strong interaction between this CTCF site and the *CUL3* promoter in our 3C data (~ 0.15 normalized promoter interaction) or in publicly available CTCF chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) (fig. S13), the sgRNA cut site is located in the middle of the predicted CTCF binding motif (JASPAR relative score: 86%). Deep sequencing of the site found mutations in 96% of alleles with a mean indel size (-9.5 bp ± 13.7 bp) that is comparable in size to the canonical CTCF motif. Using ChIP-ddPCR, we found that CTCF occupancy at this site is decreased by 45% after editing and there is a 30% decrease in *CUL3* expression (Fig. 4G). We also explored two putative AP-1 sgRNA target sites that confer drug resistance (Fig. 4D). AP-1 is a heterodimeric basic leucine zipper transcription factor, composed of FOS and JUN subunits, and its over-activation promotes metastasis in carcinomas, breast cancer, and melanoma (*41*). After editing at distal 3′ sg1 and sg2, we found decreased FOS and JUN binding compared with control cells. Editing at either site resulted in an ~25% decrease in *CUL3* expression (Fig. 4H). In keeping with observations from the global screen data (fig. S4C), mutation of these 3′ noncoding sites does not have as strong of an effect on gene regulation and function as mutations in the 5′ noncoding region.

SUPPLEMENTARY FIGURES

Figure S1

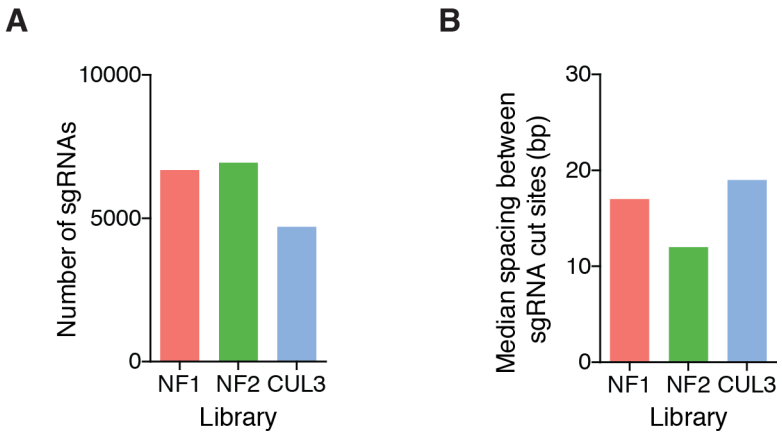

Fig. S1: Statistics of library design, sgRNA cut sites, and the locations of enriched sgRNA target sites after vemurafenib treatment in libraries targeting genomic regions near *NF1*, *NF2*, and *CUL3*.

(A) Total number of single guide RNAs (sgRNAs) in each of the 3 gene-specific libraries. **(B)** Median distance between consecutive sgRNAs (in bp) in each of the 3 libraries.

Figure S2

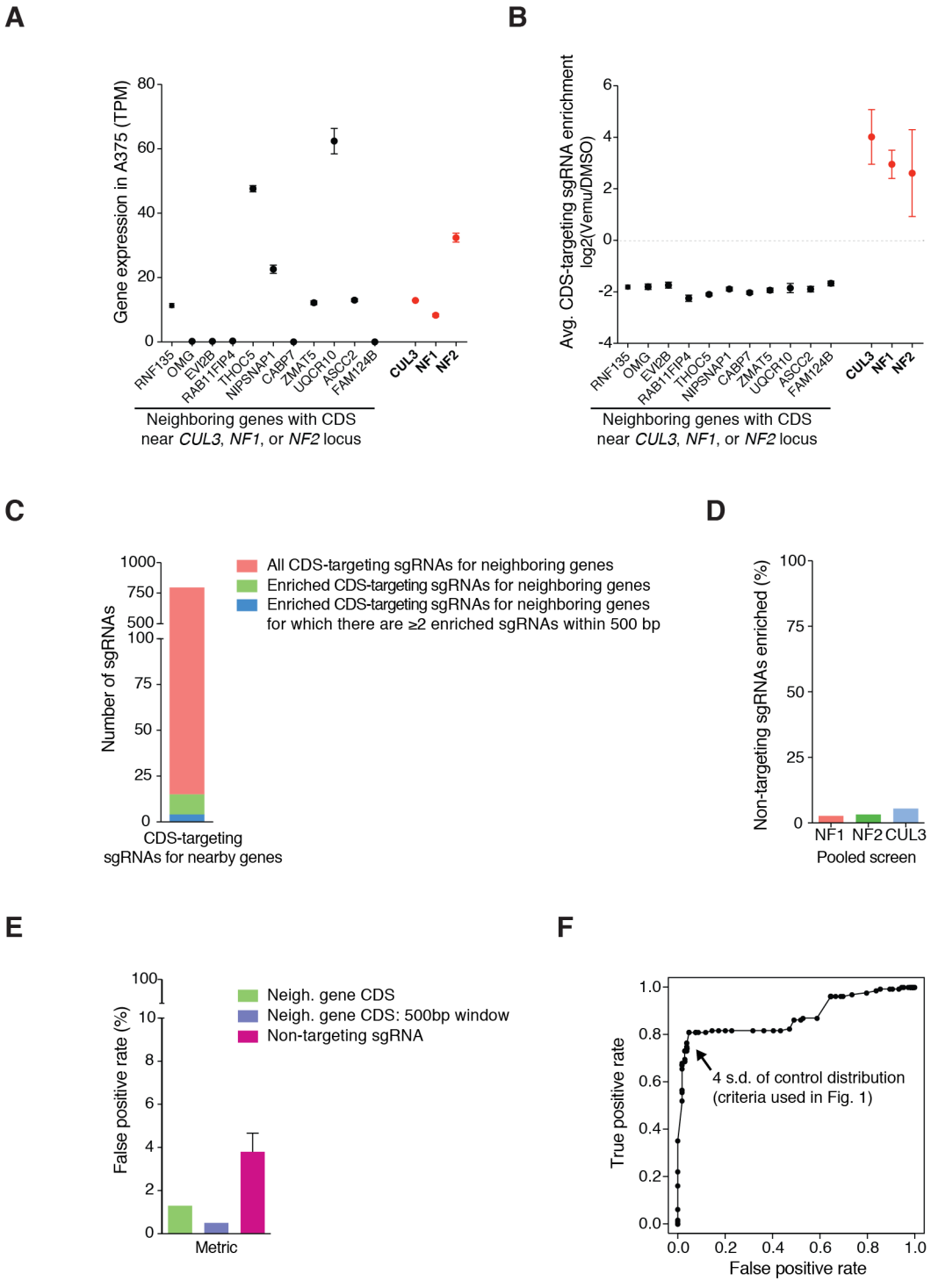

Fig. S2: Estimation of false positive rate and selection criteria for enrichment cutoff value.

(A) Expression data (transcripts per million, TPM) from A375 RNA-seq of nearby genes with coding exons that overlap regions targeted in the screen. For comparison, expression data is also shown for vemurafenib resistance genes *CUL3*, *NF1*, and *NF2*. **(B)** Average enrichment of sgRNAs targeting coding sequences (CDS) of nearby genes and of vemurafenib resistance genes *CUL3*, *NF1*, and *NF2*. For all nearby genes, the average enrichment is negative (depletion). **(C)** The number of sgRNAs over all 3 libraries that target CDS of nearby genes and a quantification of the fraction of these sgRNAs that are enriched (> 4 s.d. from control cells) and the fraction that are enriched with at least 1 other enriched sgRNA within 500 bp. Assuming that any enrichment of sgRNAs targeting CDS of nearby genes is a false positive, we can estimate a false positive rate for any enrichment (11 out of 782 = 1.4%) and for multiple enrichment within a 500 bp window (4 out of 782 = 0.5%). **(D)** Estimation of false positive rate by quantifying the percent of non-targeting sgRNAs (i.e. sgRNAs designed to not target anywhere in the human genome) that are enriched > 4 s.d. from control cells. **(E)** Comparison of false positive rates estimated from enrichment of CDS-targeting sgRNAs for nearby genes (C) and from non-targeting sgRNAs (D). **(F)** The average (over *NF1*, *NF2* and *CUL3* screens) receiver-operator characteristic (ROC) plot as a function of the enrichment cutoff (number of standard deviations from the distribution of sgRNAs in the control cells). In this plot, the true positive rate is estimated from the fraction of enriched CDS-targeting sgRNAs for genes known to confer vemurafenib resistance (*NF1*, *NF2* and *CUL3*) and the false positive rate is the fraction of enriched non-targeting sgRNAs. The point corresponding to the cutoff value used for the screen analysis (enriched sgRNAs as >4 standard deviations from the control sgRNA distribution) is denoted by the arrow.

Figure S3

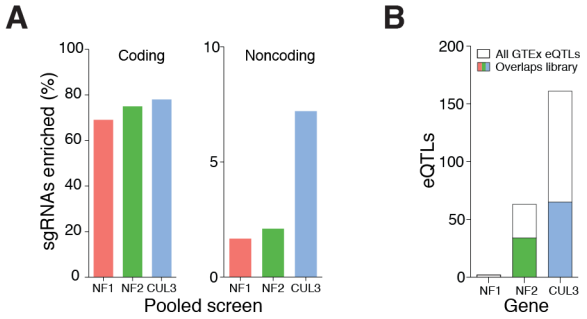

Fig. S3: Enriched sgRNAs by region targeted and overlap between all library sgRNAs and human expression quantitative trait loci (eQTLs).

(A) Percent of sgRNAs that are enriched (>4 s.d. from control cells) with target sites in coding regions (*left*) or noncoding regions (*right*) for the *NF1*, *NF2*, and *CUL3* pooled screens. Note the different scale for the coding and noncoding plots. **(B)** Total expression quantitative trait loci (eQTLs) found in the Genotype-Tissue Expression (GTEx) v6 analysis release (7,051 tissue samples from 449 donors) for *NF1*, *NF2*, and *CUL3*. Shaded regions indicate eQTLs that are contained within the region targeted by each sgRNA library.

Figure S4

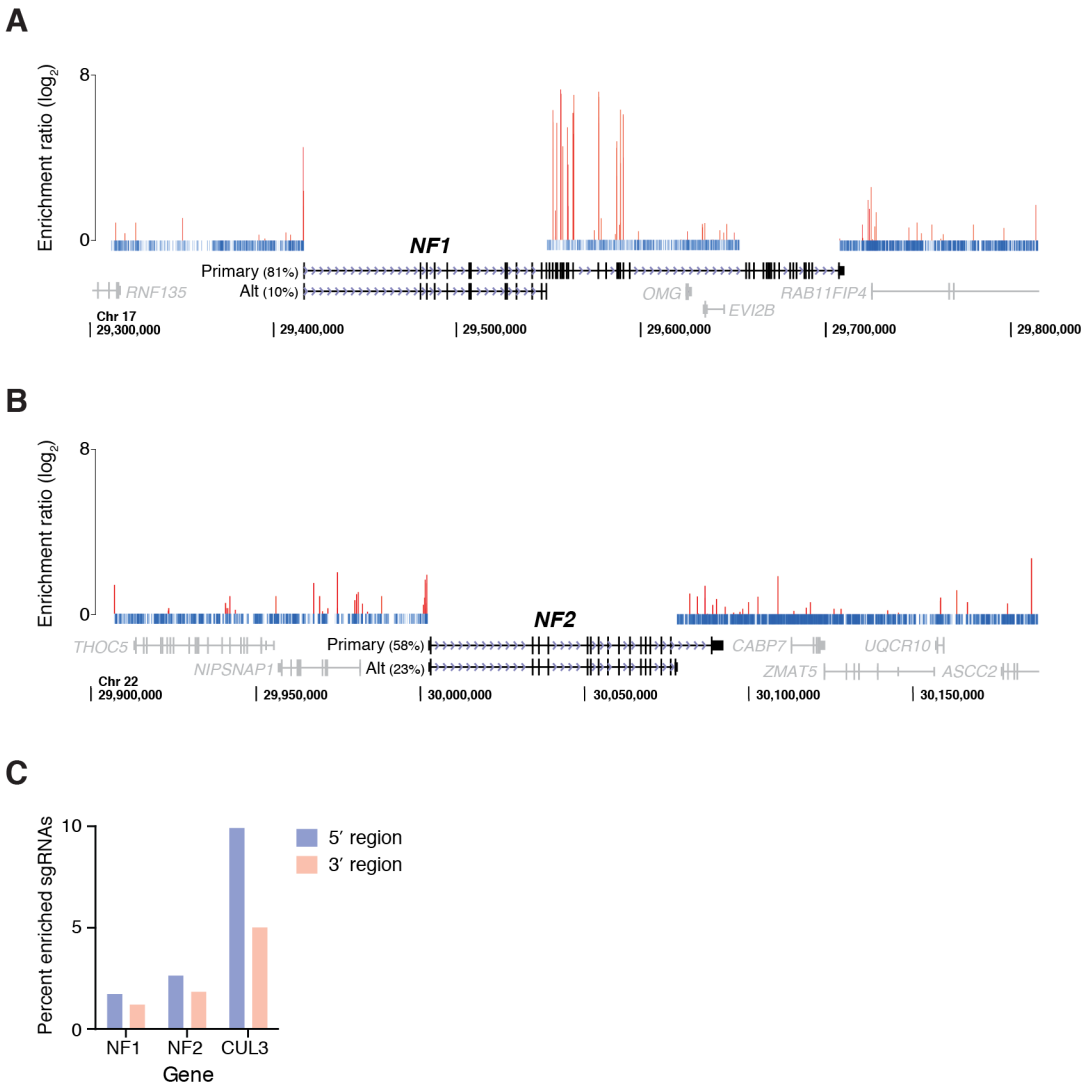

Fig. S4: Overlay of sgRNA enrichment ratio in *NF1* and *NF2* screens on chromosomal region surrounding each gene.

(A) - (B) Enrichment ratio for 2 separate mutagenesis screens targeting ~200 kb near gene loci (**A**: *NF1*, **B**: *NF2*) in A375 *BRAF* mutant cells. sgRNAs are plotted by genome coordinates (hg19) of their target site. The enrichment ratio is the log2 ratio of the normalized read count for each sgRNA in vemurafenib to its normalized read count in control (minimum value of the 2 infection replicates). Enriched sgRNAs are plotted in red with their enrichment ratio. For depleted sgRNAs (blue), only position is shown. Relative expression from RNA-seq in A375 of

the top two most highly expressed RefSeq isoforms for each gene is indicated next to the corresponding transcript. All gene-specific libraries were designed to target the proximal 100 kb from the start/end of each RefSeq isoform's coding sequence. **(C)** Each library targets ~100 kb on both 5′ and 3′ sides of the gene. In all 3 libraries, after vemurafenib treatment, there are more enriched sgRNAs (>4 standard deviations from the mean of the control/DMSO distribution) that target regions on the 5′ side than on the 3′ side of the gene.

Figure S5

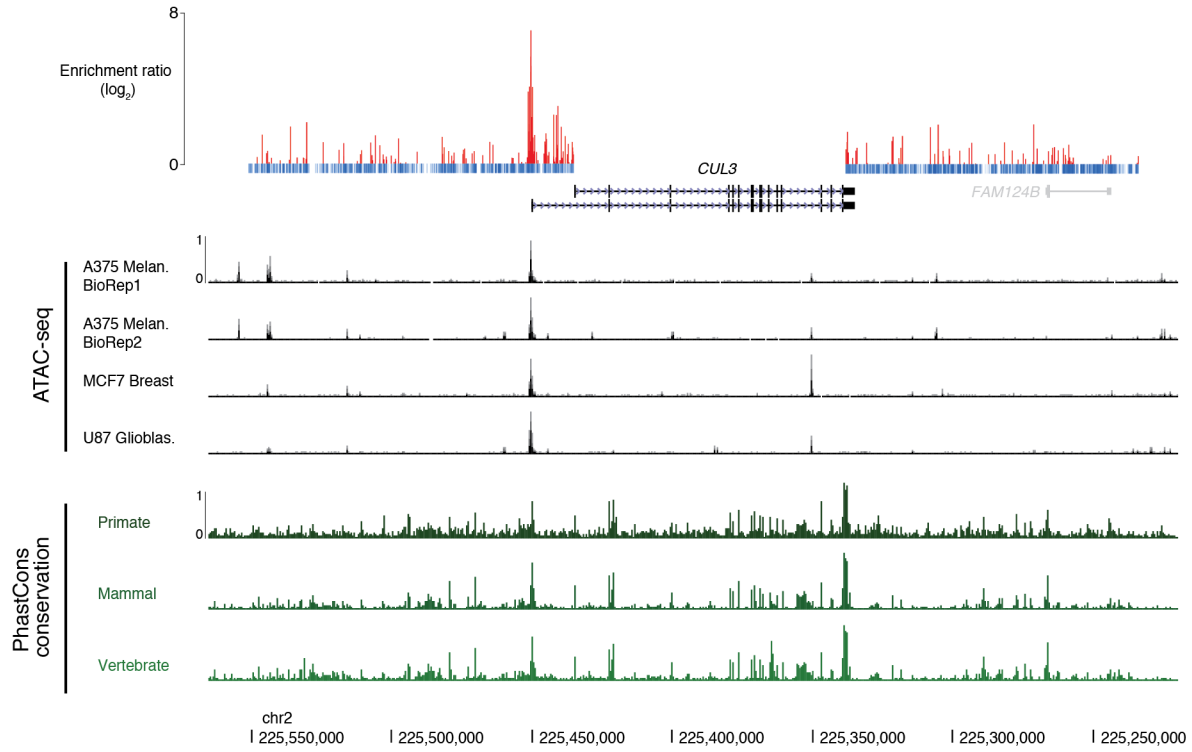

Fig. S5: Assay for Transposable and Accessible Chromatin sequencing (ATAC-seq) from 3 human cancer cell lines and phastCons conservation probabilities over the entire region targeted by the *CUL3* CRISPR library.

ATAC-seq analysis (normalized read counts) of chromatin accessibility in 3 human cancer cell lines: A375 V600E melanoma, MCF7 breast cancer, and U87 glioblastoma. Peaks indicate regions with more open chromatin. phastCons conservation scores from a phylogenetic Hidden Markov Model (HMM) trained on data from primate, mammalian, and vertebrate genomes. Higher phastCons probabilities indicate regions that are more conserved within the indicated group of organisms. The topmost track (enrichment ratio) shows the $\log_2$ (Vemu/Control) ratio for each sgRNA. Values are the minimum from 2 independent infections replicates. For clarity, only enrichment values for enriched (>0) sgRNAs are plotted (*red*); depleted sgRNAs are indicated by a short bar (*blue*).

Figure S6

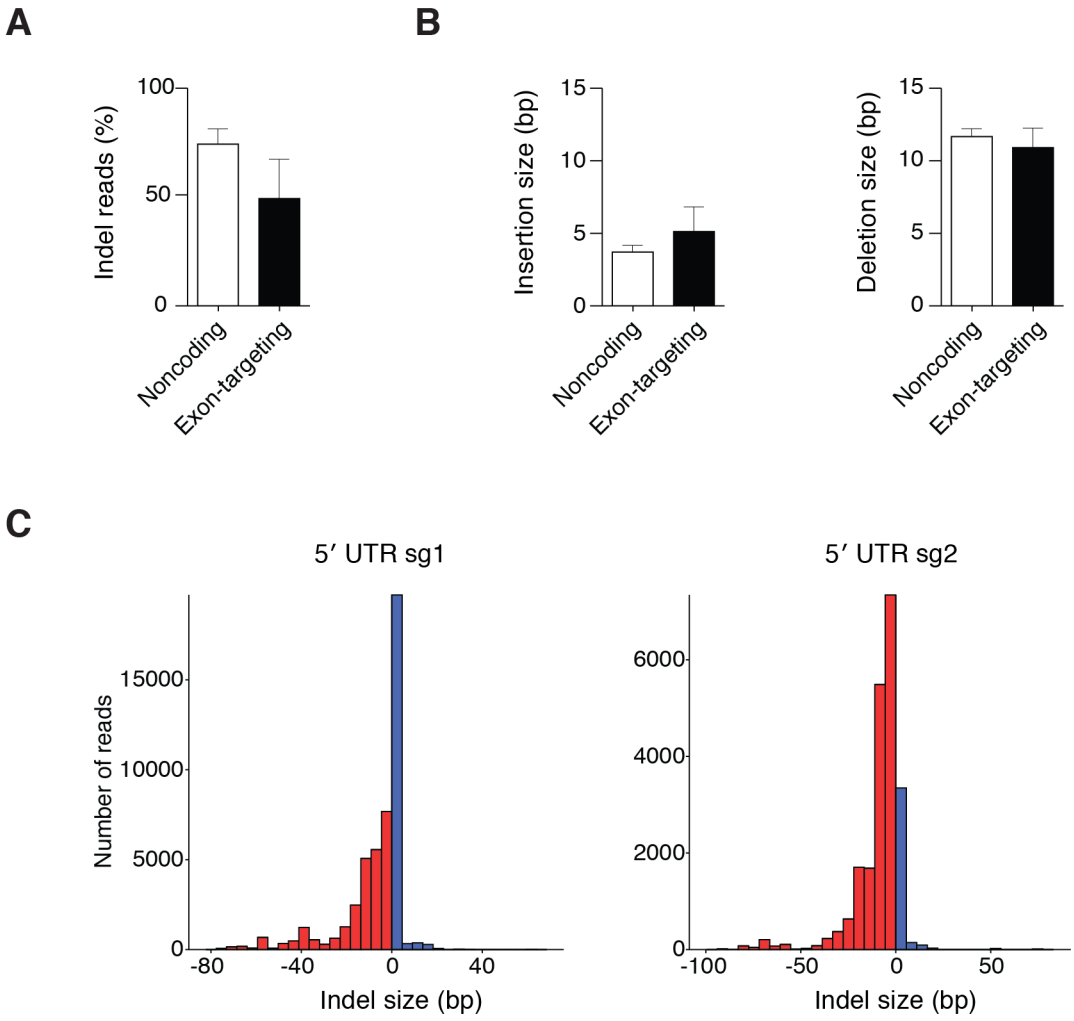

Fig. S6: Deep-sequencing analysis of insertion-deletion (indel) mutations after genome modification using validation set sgRNAs.

(A) Mean and standard error of the percent of reads containing an indel mutation for sgRNAs targeting noncoding regions near *CUL3* and coding exons of *CUL3* (*n = 24* noncoding sgRNAs, 4 exon-targeting sgRNAs). Cells were selected for lentiviral CRISPR constructs using puromycin for 7 days and then plated in R10+DMSO for a further 4 days. **(B)** Average size of insertions (*left*) and deletions (*right*) in sgRNAs targeting noncoding regions near *CUL3* and *CUL3* exons. **(C)** Histograms of indel mutation sizes for 2 sgRNAs that target noncoding regions near *CUL3*. Deletions are shown in red and insertions are shown in blue. The larger deletion size (shown in aggregate in (B)) can also be seen for these 2 sgRNAs.

Figure S7

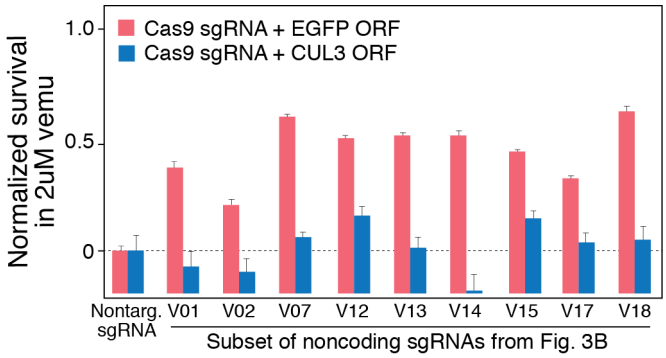

Fig. S7: Vemurafenib resistance after expression of *CUL3* open reading frame (ORF) transgene in A375 cells edited with different noncoding validation set sgRNAs.

Relative survival after 2uM vemurafenib treatment in cells transduced with noncoding sgRNAs (*red bars*) and in cells transduced with noncoding sgRNAs and a lentiviral rescue construct ("CUL3 ORF") to restore *CUL3* expression (*blue bars*). Vemurafenib survival is normalized to survival of A375 cells transduced with a *CUL3* exon-targeting sgRNA with the baseline set to the survival of A375 cells transduced with a non-targeting sgRNA. For rescue experiments, cells were first selected with puromycin only (for Cas9-sgRNA in lentiCRISPRv2) for 7 days, then transduced with the EFS-CUL3-2A-Blast lentivirus, and then selected with blasticidin for 7 days before plating for the vemurafenib resistance assay. Cas9 sgRNA cells went through the same procedure but with an EFS-EGFP-2A-Blast lentivirus substituted in place of the *CUL3* rescue construct. See Table S4 for the guide sequence and the target site location for each sgRNA.

Figure S8

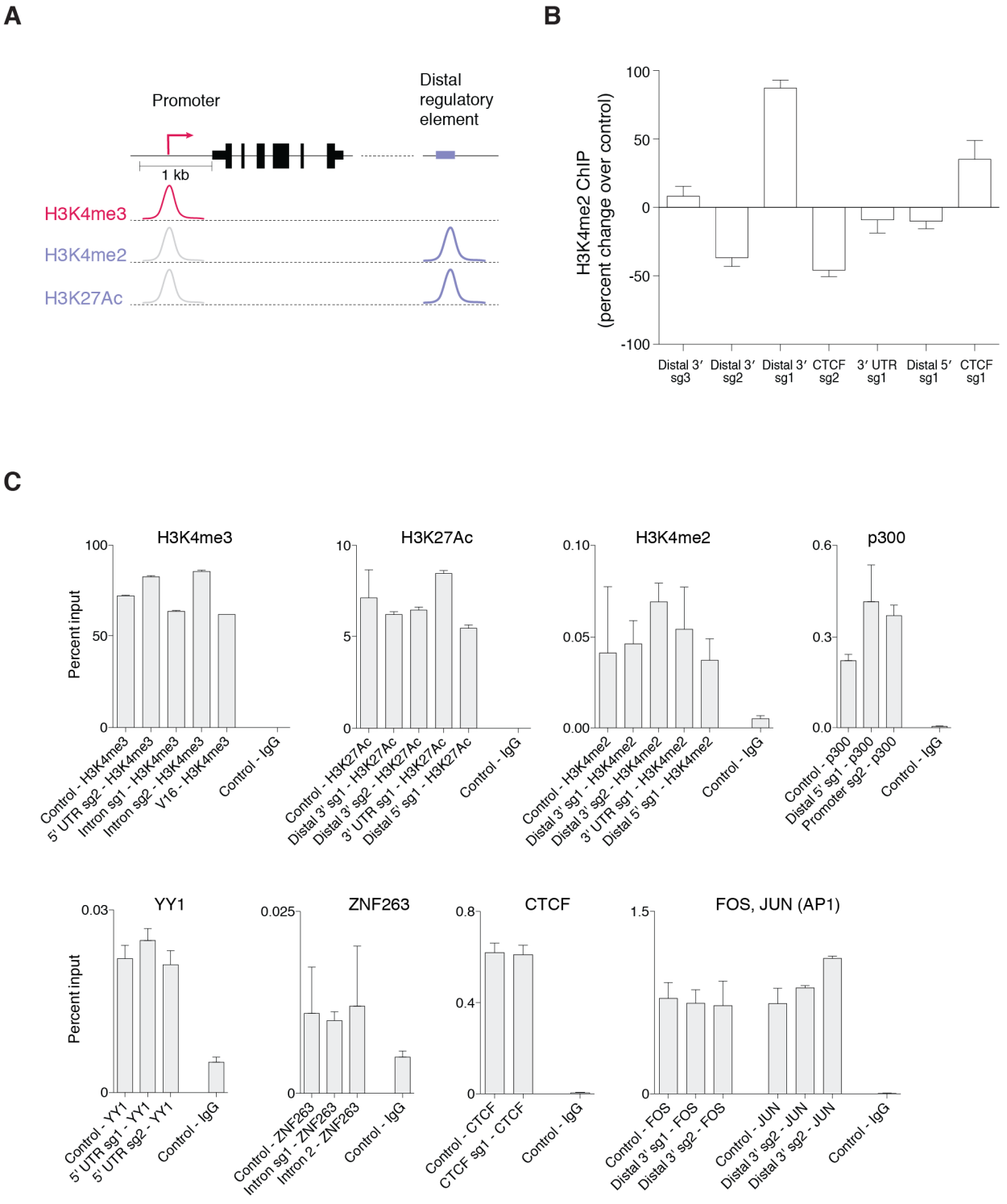

Fig. S8: Chromatin immunoprecipitation (ChIP) for individual sgRNAs for H3K4me2 and for positive control regions for all ChIP antibodies used.

(A) Schematic of histone modifications typically found at promoter proximal and distal regulatory elements. H3K4me3 is often found at the transcription start site of active or poised genes, whereas H3K27ac and H3K4me2 are found both at promoters and distal regulatory elements. **(B)** Percent change in ChIP signal (as measured by ddPCR quantification) for the H3K4me2 histone modification after genome editing by the indicated validation sgRNA. A subset of sites shows a decrease in H3K4me2 after genome editing at the site but, across all sites, there is not a significant, consistent change (p = 0.82, *two-sided t-test*). **(C)** Percent input for transcription factors and histone post-translational modifications in wild-type A375 cells and after transduction with different validation sgRNAs. In positive control regions (distant from the *CUL3* locus), the percent input is comparable between wild-type A375 and A375 transduced with validation sgRNAs. Pulldown with antibody to IgG does not result in similar levels of enrichment at any of the positive control regions. Sample labeling on the x-axis is written as *[Genome modification/Control] - [Antibody]*. The variability in percent input between different ChIP targets are due to genomic abundance (e.g. transcription factors are less abundant that histones) and differences in pulldown efficiency between antibodies.

Figure S9

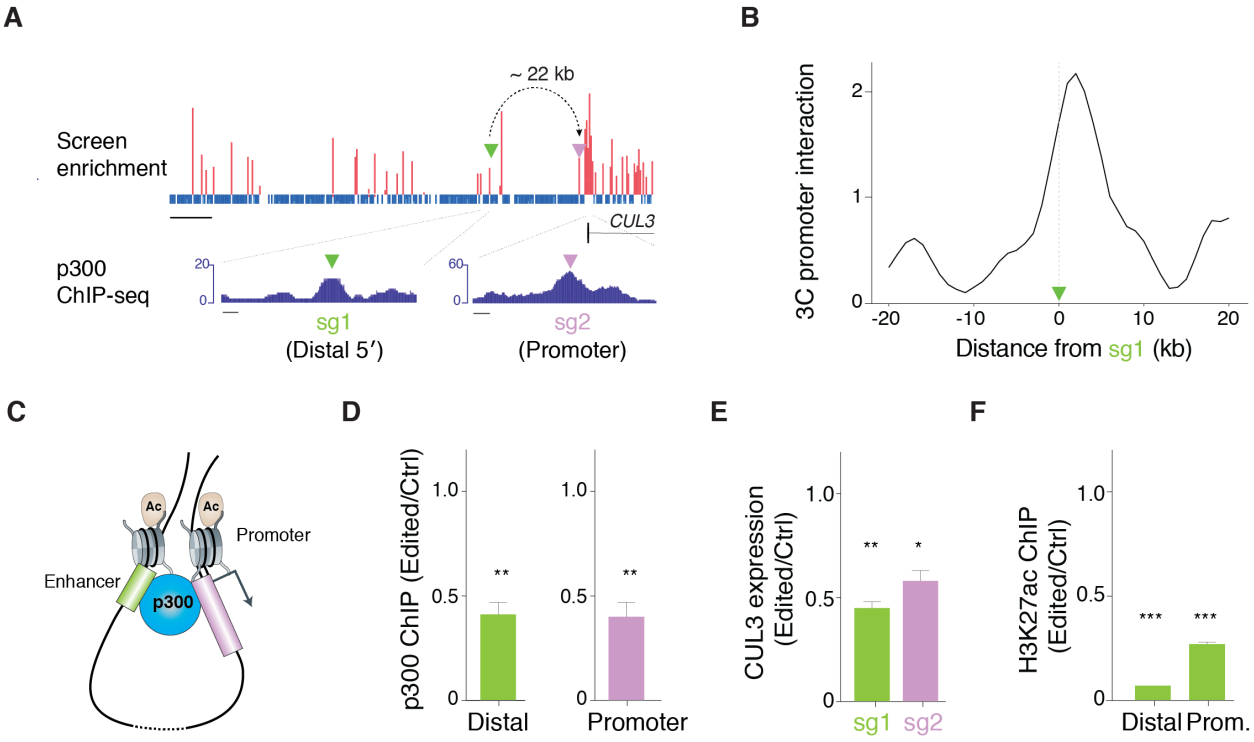

Fig. S9: Noncoding mutations at a distal p300 binding site impact CUL3 expression via long-range looping interactions.

(A) Screen enrichment near a promoter proximal and a distal sgRNA site that coincide with p300 ChIP-seq peaks (ENCODE/SYDH/p300). Dashed arrow indicates a strong interaction frequency measured between the distal site and the *CUL3* promoter by 3C. Scale bars: 10 kb (screen enrichment), 250 bp (p300 ChIP-seq). **(B)** Smoothed 3C signal measuring *CUL3* promoter interaction around distal sgRNA site in (A). **(C)** Model of chromatin looping interaction to bring p300 enhancer element into proximity with the *CUL3* promoter. **(D)** p300 ChIP around cut sites at 7 days post-transduction with distal element-targeting or promoter-targeting sgRNA (normalized to cells transduced with non-targeting sgRNA). **(E)** H3K27ac ChIP at promoter-proximal and distal sites at 7 days post-transduction with distal element-targeting sgRNA (normalized to cells transduced with a non-targeting sgRNA). **(F)** CUL3 expression at 7 days post-transduction with distal element- and promoter-targeting sgRNA (normalized to cells transduced with non-targeting sgRNAs).

Figure S10

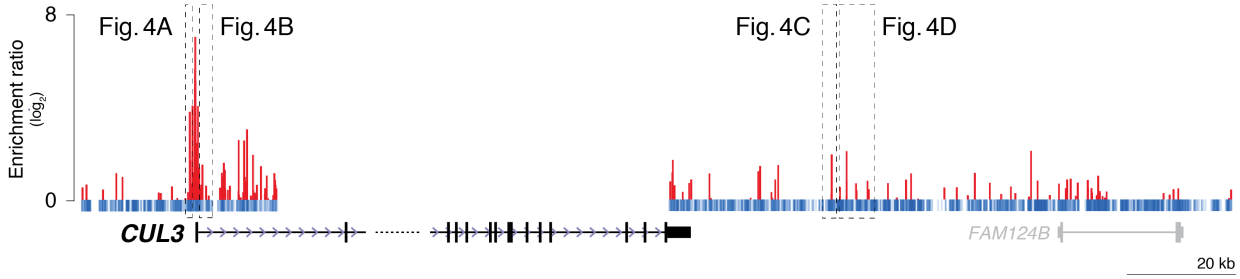

Fig. S10: Locations of predicted disrupted transcription factor binding sites.

Regions from Figure 4 with predicted transcription factor binding sites (TFBS). Location and noncoding screen enrichment of selected sgRNA target sites in the 5'-UTR (Fig. 4A), first intron (Fig. 4B) and 3´ distal sites (Fig. 4C, D) for transcription factor binding analysis. They are shown in relationship to the *CUL3* gene's 5´ and 3´ regions with the noncoding mutagenesis screen enrichment plotted above. Scale bar: 20 kb.

Figure S11

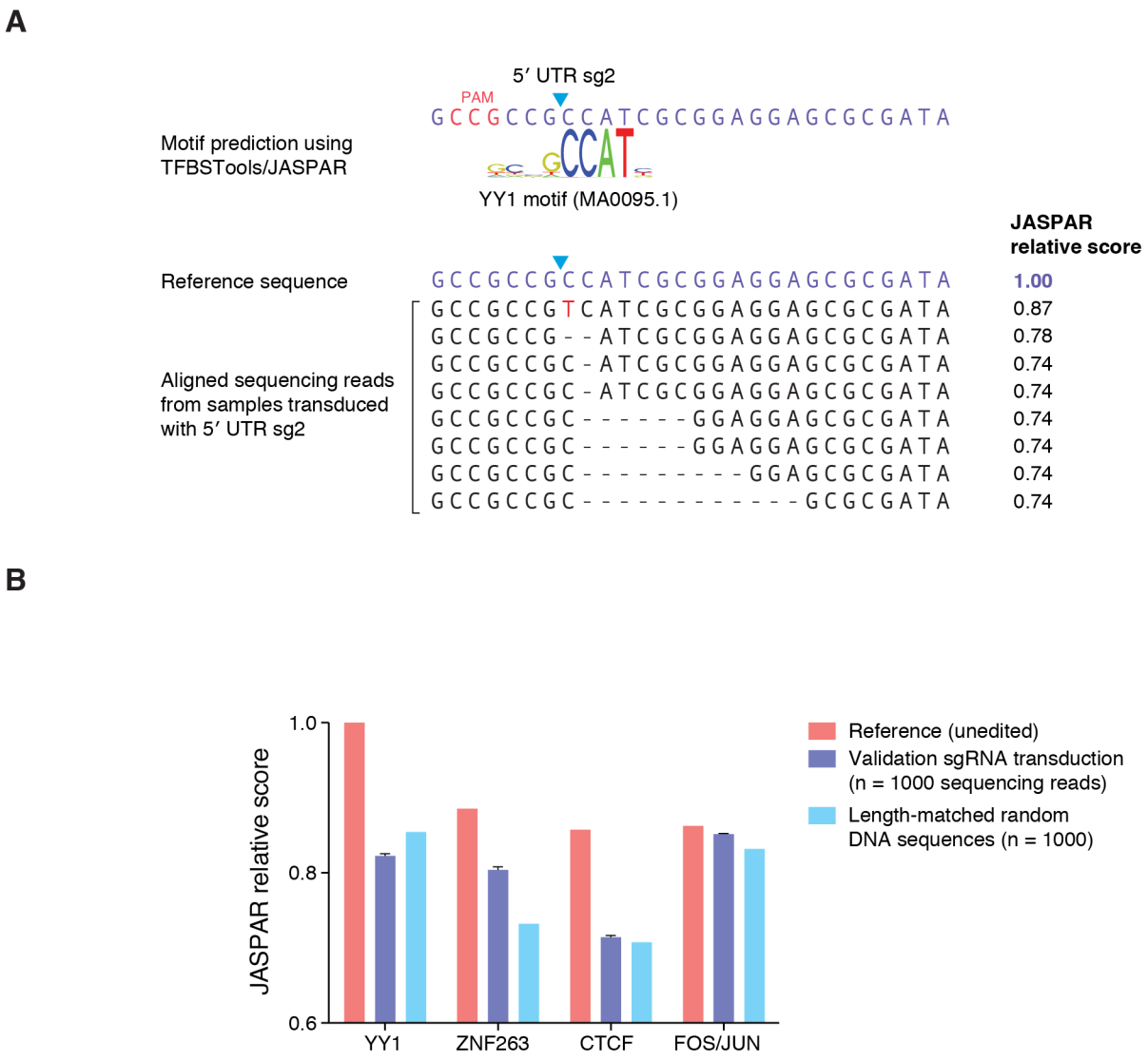

Fig. S11: Deep sequencing of indel mutations after genome modification to bioinformatically predict disrupted transcription factor (TF) binding sites.

(A) An example of a predicted TF binding motif for one validation sgRNA. In this case, JASPAR relative scores for the TF binding (using the indicated position-weight matrix from the JASPAR database) were computed both for the genome reference sequence (*hg19*) and sequences from cells transduced with a validation sgRNA (*5′ UTR sg2*) after 7 days of puromycin selection (followed by 4 additional days of cell culture with R10+DMSO). A JASPAR relative score of 1 (as scored by the reference sequence) is defined as the maximum

likelihood sequence for the motif. That is, the most probable motif base at each position is found in the tested sequence. Sequences with various indel mutations near the sgRNA cut site (blue arrow) have different (and, in this case, lower) JASPAR relative scores, implying that the TF binding site may have altered affinity for the TF after genome modification. (B) Comparison of JASPAR relative scores for the indicated TF before (red bars) and after (purple bars) genome modification. Relative scores before genome modification were computed using the reference sequence (as in (A)). Relative scores after genome modification were computed by random sampling of 1,000 sequencing reads containing indels after genome modification by the corresponding validation sgRNA and computing the average JASPAR relative score (error bars are standard error). Validation sgRNAs and JASPAR motifs used were: 5′ UTR sg2 (YY1, MA0095.1), intron sg2 (ZNF263, MA0528.1 modified to match DeepBind motif (*42*)), CTCF sg1 (CTCF, MA0139.1), Distal 3′ sg1 (Jun/Fos, MA0099.2). We also generated random DNA sequences the same length as the indel reads to estimate a background binding rate (assuming a randomly distribution of nucleotides) for each TF motif. This is useful because some motifs are quite short and thus high-scoring binding sites can occur by chance frequently. We then computed JASPAR relative scores for these 1,000 length-matched random DNA sequences. In all cases, the reference sequence provided the best match (highest JASPAR relative score) for the TF shown and, in all cases, the average relative score was lower after genome modification. In many of the cases, there was no significant difference between the JASPAR relative score after genome modification and relative scores computed from length-matched random DNA sequences, suggesting a complete loss of the motif.

Figure S12

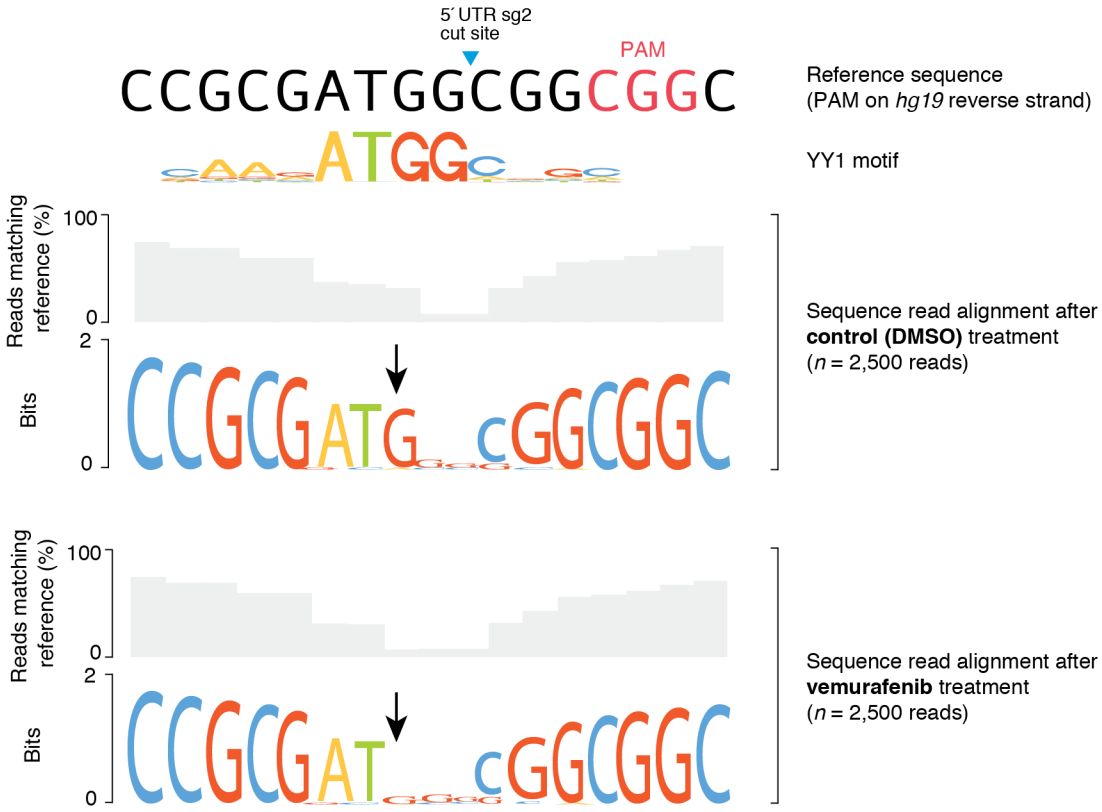

Fig. S12: Vemurafenib treatment selects for YY1 motif-damaging indel mutations.
Multiple sequence alignment (iterative *k*-mer aligner from Geneious R6) of 2,500 sequencing reads from A375 cells transduced with an sgRNA from the validation set (*5′ UTR sg2*) and selected with puromycin for 7 days. After selection, cells were replaced in either R10+vemurafenib or R10+DMSO (control) and grown for 4 days before extracting genomic DNA and preparing libraries for sequencing. Compared to the control treatment, A375 cells treated with vemurafenib have more indel mutations that damage a YY1 binding motif. After vemurafenib, there is a decrease in the number of reads matching the reference sequence at the indicated base (black arrow) and an increase in entropy (as measured by information content in bits) at the indicated base.

Figure S13

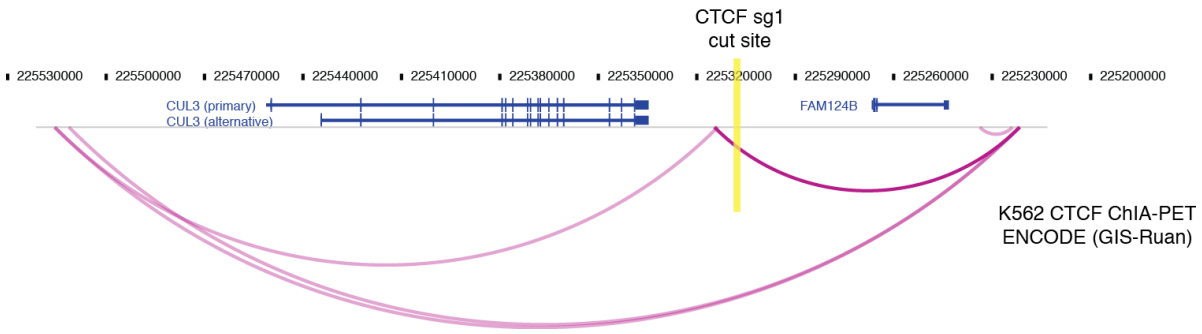

Fig. S13: CTCF sg1 targets a CTCF site without a strong direct interaction with the *CUL3* promoter.

In a publicly available CTCF chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) dataset from K562 cells (ENCODE/GIS-Ruan), there does not appear to be a strong interaction between the region targeted by CTCF sg 1 (*yellow box*) and the *CUL3* promoter. There is some evidence for interaction at a nearby site (<10 kb away) with the promoter.

SUPPLEMENTARY TABLES

Table S1: Overlap between ATAC-seq and DNAseI HS-seq peaks and noncoding screen functional enrichment

Table S2: Pooled mutagenesis screen Illumina sequencing primers

Table S3: Chromosome conformation capture (3C) enzyme cut sites and primers

Table S4: List of validation sgRNAs and target sites

Table S5: Genomic and barcode primers for targeted indel sequencing

Table S6: Chromatin immunoprecipitation-droplet digital PCR (ChIP-ddPCR) primers

Table S7: ChIP antibodies and optimized concentrations

Table S1: Overlap between ATAC-seq and DNAseI HS-seq peaks and noncoding screen functional enrichment

	Peaks identified	Enriched sgRNAs within 100 bp of ATAC/DNAseI peaks	ATAC/DNAseI peaks within 100bp of enriched sgRNAs
ATAC-seq (A375_melanoma)	11 peaks	7% of enriched sgRNAs	8 of 11 ATAC peaks
DNAseI HS-seq (Colo829_melanoma)	5 peaks	4% of enriched sgRNAs	3 of 5 DNAseI HS peaks

Table S2: Pooled mutagenesis screen Illumina sequencing primers

Name	Sequence
PCR1_F	AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG
PCR1_R	CTTTAGTTTGTATGTCTGTTGCTATTATGTCTACTATTCTTTCC
PCR2_F01	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTtAAGTAG AGtcttgtggaaaggacgaaacaccg
PCR2_F02	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTatACACG ATCtcttgtggaaaggacgaaacaccg
PCR2_F03	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTgatCGCG CGGTtcttgtggaaaggacgaaacaccg
PCR2_F04	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTcgatCAT GATCGtcttgtggaaaggacgaaacaccg
PCR2_F05	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTtcgatCG TTACCAtcttgtggaaaggacgaaacaccg
PCR2_F06	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTatcgatT CCTTGGTtcttgtggaaaggacgaaacaccg
PCR2_F07	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTgatcgat AACGCATTtcttgtggaaaggacgaaacaccg
PCR2_F08	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTcgatcga tACAGGTATtcttgtggaaaggacgaaacaccg
PCR2_F09	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTacgatcg atAGGTAAGGtcttgtggaaaggacgaaacaccg
PCR2_F10	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTtAACAAT GGtcttgtggaaaggacgaaacaccg
PCR2_F11	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTatACTGT ATCtcttgtggaaaggacgaaacaccg
PCR2_F12	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTgatAGGT CGCAtcttgtggaaaggacgaaacaccg
PCR2_R01	CAAGCAGAAGACGGCATACGAGATAAGTAGAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC TtTCTACTATTCTTTCCCCTGCACTGT
PCR2_R02	CAAGCAGAAGACGGCATACGAGATACACGATCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC TatTCTACTATTCTTTCCCCTGCACTGT
PCR2_R03	CAAGCAGAAGACGGCATACGAGATCGCGCGGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC TgatTCTACTATTCTTTCCCCTGCACTGT
PCR2_R04	CAAGCAGAAGACGGCATACGAGATCATGATCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC TcgatTCTACTATTCTTTCCCCTGCACTGT
PCR2_R05	CAAGCAGAAGACGGCATACGAGATCGTTACCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC TtcgatTCTACTATTCTTTCCCCTGCACTGT
PCR2_R06	CAAGCAGAAGACGGCATACGAGATTCCTTGGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC TatcgatTCTACTATTCTTTCCCCTGCACTGT
PCR2_R07	CAAGCAGAAGACGGCATACGAGATAACGCATTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC TgatcgatTCTACTATTCTTTCCCCTGCACTGT
PCR2_R08	CAAGCAGAAGACGGCATACGAGATACAGGTATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC TcgatcgatTCTACTATTCTTTCCCCTGCACTGT
PCR2_R09	CAAGCAGAAGACGGCATACGAGATAGGTAAGGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC TacgatcgatTCTACTATTCTTTCCCCTGCACTGT
PCR2_R10	CAAGCAGAAGACGGCATACGAGATAACAATGGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC TtTCTACTATTCTTTCCCCTGCACTGT
PCR2_R11	CAAGCAGAAGACGGCATACGAGATACTGTATCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC TatTCTACTATTCTTTCCCCTGCACTGT
PCR2_R12	CAAGCAGAAGACGGCATACGAGATAGGTCGCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC TgatTCTACTATTCTTTCCCCTGCACTGT

Table S3: CUL3 Chromosome conformation capture (3C) enzyme cut sites and primers

Enzyme	Side	Primer sequence	Primer coordinates (hg19, chr2)	Enzyme coordinates (hg19, chr2)
BglII	**bait**	**CCTGAGCGAGACGAGAT**	225450119	225450242
BglII	left	TGGTGGGAGGTGATTGA	225235052	225235111
BglII	left	ATAGTTTGGCTGTATCCCTATG	225234985	225235111
BglII	left	TTTCTAAGTGACGTGGGTTTAG	225237511	225237570
BglII	left	GCATCTAGGCCTTCAGTTAG	225250239	225250285
BglII	left	CCTGGGAGCTCTGAGAATA	225258453	225258548
BglII	left	CTGCCACAATTCCCATGT	225261858	225261950
BglII	left	GACCCTAAGGGACGCTAATA	225265655	225265733
BglII	left	CCTGTGTCTGCAGTTTCTC	225274018	225274128
BglII	left	GCATATTCTGGTCTCCTAAGTC	225274085	225274128
BglII	left	GTCTGCCCTGCAGAATAAAG	225298258	225298356
BglII	left	TTTCTGGAGAATCCTGACTAATG	225303341	225303422
BglII	left	TTTGAGGAGGAGTTTCGCT	225312056	225312090
BglII	left	GGTGACACATGCCTGTAAT	225312885	225312888
BglII	left	TGTGCCACTCAAGACAATC	225315736	225315839
BglII	left	TGAAGAAACCATCTAAGTCATC	225317872	225317938
BglII	left	AATTAGCTGGGCATGGTG	225320413	225320501
BglII	left	CCTCACAATCATGGCAGAAG	225322279	225322374
BglII	left	AGAAACACTGCATCATCTAGG	225332147	225332241
BglII	left	CCAGCAATCTCCAACCATTC	225336843	225336935
BglII	right	CGAAGGCTTCTTCCAACTC	225438282	225438352
BglII	right	TCCTCTAGCATTAGGGAGTG	225444385	225444475
BglII	right	CATTGTGGAGATCAAATGTGC	225445725	225445789
BglII	right	TCTTTCCTCACTGCAACTG	225448537	225448639
BglII	right	TTTCTGTGCCCAGTCATATTC	225453043	225453146
BglII	right	CCTCTTCTTGACCATCAGTTTC	225453380	225453424
BglII	right	TCCCATTGTGTGAACCTAAC	225456453	225456557
BglII	right	GTACTATGGGTAGGAAACTGTTC	225460687	225460810
BglII	right	CGCTTGACCCTGTCTTTAC	225462338	225462439
BglII	right	AGAGACGGAGACACACATAG	225473151	225473257
BglII	right	GTTGAAAGAAGGCAACTAGAATAAG	225478405	225478486
BglII	right	CAGTGATACACACACAGACAC	225491758	225491920
BglII	right	GGGATCTAAATGAGAGGATCAC	225509977	225510067
BglII	right	TTCTTCTGCCAGATACCTAAATC	225527904	225527934
BglII	right	TGGGAGGCCTCAGAATC	225528355	225528448
BglII	right	ATCGTGCCACTGCACTC	225538300	225538470
BglII	right	TAGCATAGTGTGTTCAAGGTTC	225538680	225538812
BglII	right	GTGAGCAGATCAAACGATTATG	225540095	225540204
BglII	right	CTTACGATCATGGCAGAAGG	225540360	225540466
BglII	right	GGCTACAGCCTTGGTATTG	225543376	225543454
BglII	right	GGGACACATGCAATTATTGAG	225545398	225545433
BglII	right	TCTGGTTTAGCATGGCTTATAG	225546972	225547061
BglII	**ctrl**	**CTTCCTTCAGTTCCCTGTTC**	225450347	225450242
HindIII	**bait**	**ACAGCTCTCAGGACTGGAAGGTG**	225450817	225450856
HindIII	left	CCTGCTCCACCCTCAAATCTCACATC	225238223	225238265
HindIII	left	GCCTATACAGGCATACCTTGTTTTATTG	225238415	225238479
HindIII	left	CATTGGAAGAAGATGCCATCTAGGAC	225239728	225239765
HindIII	left	GCCAAAATAAGTCTGCCTGGGTTCAG	225244542	225244590
HindIII	left	GCATCTAGGCCTTCAGTTAGCGTC	225250243	225250298
HindIII	left	CTTCTGTGTGGGATGTGCATCCTCTAG	225251015	225251056
HindIII	left	GTATGTCCAGTGCCTAGCACAGTG	225251461	225251518
HindIII	left	CAATTCTATGTGCTATATTCTTTAAAACTGTAATGG	225256337	225256408
HindIII	left	GGCAACAGACCAAGACTCTGTCTC	225257470	225257549
HindIII	left	CACCTGTTTGAGACACCCTTGCTC	225261045	225261118
HindIII	left	GCCTTTACACACTTTCCTCAGGCAC	225263223	225263289
HindIII	left	GTTTCCTAGTTATTGTGAGCAGCTCAG	225268958	225269034
HindIII	left	GGCTCCTTCTAGGGCAGAGGTG	225272032	225272096

HindIII	left	GAGGCTCAAAGAAGGGTATGAGAC	225273516	225273571
HindIII	left	CATGGCACCTGTAGCAAATGCTAGAC	225275607	225275666
HindIII	left	CTGAGACTGTGGTTTCTATGGCTG	225277244	225277303
HindIII	left	GAGCTGGGAGGGAATTGCATACC	225279524	225279548
HindIII	left	GCTCTTAAGAGGTCTAAGAAGAAACTTCC	225282405	225282465
HindIII	left	CTTCCATAGATGCTTACCCAGTGG	225283454	225283502
HindIII	left	GCACTGATGCAAAGGAATGCTCTGG	225283670	225283718
HindIII	left	GGTTTCTCTTCTGGTGAACTTCAAACAC	225297233	225297289
HindIII	left	GATTCCCAGTGCCTGACACATACTAG	225298673	225298686
HindIII	left	CATGGCCACAGAAGACATTCTGCC	225303645	225303721
HindIII	left	GGGTGAGCATTACATAAGCAACCTG	225305863	225305928
HindIII	left	GGTTCTATTCTGGCCCAGGTAGTCAG	225305986	225306128
HindIII	left	GACCTGGTCCATCCGGTTCTGATG	225307447	225307472
HindIII	left	CTTTGTTACAGCAGCTGGAACAGACCAAG	225313300	225313362
HindIII	left	GTTTCTGACATTTAAGTGGCATTTTGCAG	225321344	225321366
HindIII	left	GATCAGGGAAGGTGCAATGAAACC	225322604	225322644
HindIII	left	GAGAACTCACTAAGTGACAGATACCC	225331162	225331220
HindIII	left	GCTGCCCACAAGAATCACCTCAG	225332810	225332867
HindIII	left	GCTCAAGGGAAGACTGGAGAATATGG	225333349	225333445
HindIII	left	GCCTATTGCTAGAGTTGCACTGGAAC	225333625	225333678
HindIII	left	GATGACAGCCTAGGCAACACAGCAAG	225334357	225334413
HindIII	right	CAAGGGAAAATACTTGATCTTAATTTCAAGCTC	225434665	225434697
HindIII	right	CTACTTATGACATCTGCAATAATACCATTTATCC	225434971	225435027
HindIII	right	GAGTAGGCTATCCAAAACTCAATTTGAG	225435575	225435622
HindIII	right	CAACTCTTTCGACTATATCTCTGTGAATGAC	225435712	225435817
HindIII	right	GGAGCTAGAATAAGCCTAAGGTAACC	225436619	225436671
HindIII	right	GTACCATGTCAACTCAAATAATCAGAGTG	225436762	225436880
HindIII	right	CAAATGTTACTGAACAATACACATTTCCCAAG	225438193	225438248
HindIII	right	GCTTATTATGTGCCAAGCACTATTC	225440350	225440359
HindIII	right	CTCATGTAATCAATCATTCACTAACCACTC	225441310	225441358
HindIII	right	GGCCTAATCGTGGCTAAATATTGG	225441846	225441888
HindIII	right	GCTGTCCATGCTACACAAGTGGAGTTC	225444367	225444429
HindIII	right	GTGGTCCTTGTTCCTCTGCATAC	225444868	225444933
HindIII	right	GTTGACTGTAAGGTTGAATTTGCCC	225451431	225451460
HindIII	right	GCTGCGTCTAAAAGCATCACTGTGAACTG	225452687	225452703
HindIII	right	CCTGCAAGGGCCATTATCACCTGGAG	225456766	225456815
HindIII	right	GCGGTGAGTGTTACAGCTCATAAAAGCAG	225467755	225467814
HindIII	right	CATCTTAAATTCGAACTCTATTAAATGGTG	225471353	225471400
HindIII	right	GATATATTTGTATACTCATGTTCATAGAAGC	225474373	225474400
HindIII	right	GTGTATCACCTAAAGGCCTTCAGATTC	225481441	225481505
HindIII	right	CCAGGTATGATGCCATGGATCTTTGG	225482978	225483026
HindIII	right	CCAGCCTGGGCAACAAGAATGAAAC	225485107	225485167
HindIII	right	GAGATTCATCCTGGGGGATTCATGGC	225497789	225497844
HindIII	right	GTGGTGAATGGATACGCCAGTTCCAG	225501679	225501727
HindIII	right	CCGTCCTAGAATAAACATAGCCATCAG	225503620	225503676
HindIII	right	CTTTGGGCTGACTCTGTGGGAAG	225508744	225508766
HindIII	right	CCTCATCTGAAAGGCAGAGTAGTAATAATTATG	225508863	225508926
HindIII	right	GAGATCAACCATGCCTACTTGTCTCC	225519049	225519083
HindIII	right	GCAGTACTGTTTCTGTGGTTCCCAG	225529590	225529636
HindIII	right	GACACAGCTAAACCATATTAACTAGCTAC	225540629	225540671
HindIII	right	CAGAAACCACAGGGGTAAGCTCTTAAAAG	225543135	225543215
HindIII	right	CTTTTAATAGTTTGAATTCTGTTTGGCTTCTG	225547828	225547853
HindIII	right	GTGCCAAGGTTCTTTCAAGTGGTTG	225549840	225549868
HindIII	**ctrl**	**CATGAATAAGCCCTGGGTCCACCAG**	225450909	225450856
EcoRI	**bait**	**TTCTCCTAAATTCCATCGTACC**	225448349	225448415
EcoRI	left	CCCAGAACTTGGGATACAAAC	225234594	225234653
EcoRI	left	TGCTCAAGGTCACATCAATAG	225241814	225241896
EcoRI	left	CATATGGGCAACGAGAATTTG	225243160	225243294
EcoRI	left	CCCTCAGATGAACAACTAACAG	225244211	225244302
EcoRI	left	ACCTCACTGGATGTTGTAAATG	225245271	225245361
EcoRI	left	ATGTTTGGCATTGGAATGAAG	225251793	225251893

EcoRI	left	TCTCAGTACAGGGAGGTAAC	225256743	225256852
EcoRI	left	CAGGAGAAGTGGGTAAAGAAG	225258393	225258498
EcoRI	left	ATCACGCCATTGCACTC	225283346	225283493
EcoRI	left	CAGGAGGATCGCTTGAG	225284395	225285041
EcoRI	left	TCTCCTCAGAGAGACTATAAACC	225286854	225286944
EcoRI	left	CATCACTAATCATCAGGGAAATG	225288766	225288871
EcoRI	left	TAAATGCAGGCTGTGGTG	225290229	225290366
EcoRI	left	AACTGAATACACAGTGAGAAGG	225290839	225290932
EcoRI	left	TGACTAGTTATTGGGTCCTATTATG	225291051	225291152
EcoRI	left	GCATACCTCCCAAAGAGAAC	225298351	225298478
EcoRI	left	GAACCAATCTCCCACAGATAC	225304880	225304962
EcoRI	left	TGTTTGTGTAGGATGCAAAGTG	225306745	225306813
EcoRI	left	CTCAGCCTCCCAAGAAG	225310296	225310390
EcoRI	left	GTGCATGACCAAGAGAAGAC	225310572	225310651
EcoRI	left	CTTGACCTCAAGTGATCCTC	225311732	225311802
EcoRI	left	AGTATTCTCGTCTTACATATGCTG	225311856	225311996
EcoRI	left	CCATGATCCACTCTTAATTTC	225314744	225314855
EcoRI	left	AACTGTTTCTTTGCCTTTCTTC	225318893	225318948
EcoRI	left	ACCACACTCAGCCTGTTAG	225320916	225321043
EcoRI	left	CCAAGTTAAGCTTAGAGAGTACAT	225322656	225322751
EcoRI	left	CCCATTGTTTGTCTGGTATAGA	225323019	225323101
EcoRI	left	GGCATGGGCCAATAAATAGA	225325226	225325322
EcoRI	left	CCATGGTTGGACTTCCATTA	225329455	225329534
EcoRI	left	GTGCCTACATCCACTACATAC	225334816	225334867
EcoRI	left	TCTACAGTACAGATGGAGACA	225337401	225337502
EcoRI	right	CACATCTTGAAGGTTCTGTGA	225434424	225434480
EcoRI	right	GCTTGTCACTGTCCTACTATT	225435835	225435927
EcoRI	right	AACCCTTGTGAATGGGATTAG	225437427	225437520
EcoRI	right	GACAGGGCAAACAGAAGAG	225438742	225438810
EcoRI	right	GAGGGAAGGAGTCGAGAAT	225439791	225439925
EcoRI	right	AGCCTGGACACCAAGAG	225455161	225455243
EcoRI	right	GCACTTCGTAAATATCTGCTTG	225463459	225463590
EcoRI	right	TGTATCATTTATGTCAGACTCCTG	225470402	225470431
EcoRI	right	AGAAATCAAGAGGAGTATATGACC	225476429	225476444
EcoRI	right	AGACGGAAGTTGCAATGAG	225483683	225483799
EcoRI	right	TGAACTCTGACTTACCCTGAG	225494198	225494335
EcoRI	right	CAGATCATGTAGAGCCTGATG	225495947	225496039
EcoRI	right	CTCTCAAAGTGCTGGGATTAC	225496576	225496666
EcoRI	right	GAAGCAATCGTTTCATCATAGTC	225498192	225498260
EcoRI	right	CTGAACGTAACTGCCTAGC	225515099	225515192
EcoRI	**ctrl**	AGTTGCAGTGAGGAAAGAC	225448536	225448450

Table S4: List of validation sgRNAs and target sites

sgRNA	sgRNA guide sequence (5' to 3')	Target location	Name in figures
V01	GGCACTTGGAATCCACATGA	3' of *CUL3*	
V02	GCAGCGTCCGGAGTTGGCAC	3' of *CUL3*	Fig 4E: Distal 3' sg2
V03	AGCACACAGTCATAACCACA	3' of *CUL3*	Fig 4E: Distal 3' sg1
V04	GCCACAGCCATGCCCAGTGG	3' of *CUL3*	Fig 4D: CTCF sg1
V05	ACTGGCTGGAATCTGCCAAG	3' of *CUL3*	
V06	CTGATCTTGAGTTGGTCCTT	3' of *CUL3*	
V07	TTAGGGGCAGGGAGGACCTA	3' UTR	
V08	GGAAATCTCAAATTACAACA	3' UTR	
V09	AAATGTACTGTTAACGAACT	Intron and promoter	
V10	AGTATATAGGATATAACTTT	Intron and promoter	
V11	CAAGAGTTTGTAAAGTGCTT	Intron and promoter	
V12	TCCGCGGCTGCTAGCAGCGC	Intron and promoter	Fig 4C: Intron sg1
V13	CGCGGAGTCCTCCCTGTGTG	Intron and promoter	Fig 4C: Intron sg2
V14	GCGCTCCTCCGCGATGGCGG	5' UTR	Fig 4B: 5' UTR sg2
V15	AGGAGGAGGAGGACGACGTT	5' UTR	Fig 4B: 5' UTR sg1
V16	AGGGGGGAAGTTCGGAGAGC	Intron and promoter	
V17	ATAGTCTTGAGGAGGAGCGT	Intron and promoter	Fig S9: Promoter sg2
V18	AAAAACACAGGAACCAGTTC	Intron and promoter	
V19	ATCTTTGTCTGACTACCTGC	5' of *CUL3*	Fig S9: Distal 5' sg1
V20	AATTTGGCTCGTCCAAACTG	5' of *CUL3*	
V21	ACAGCTTCTACTCTTAGGTC	5' of *CUL3*	
V22	GATATAGTGAAGTCCAACAA	5' of *CUL3*	
V23	TGTAGGAGAATGTGCAAGGA	5' of *CUL3*	
V24	CACACACTCAGATGGCTACA	5' of *CUL3*	
V25	GTTAGAGCACCAGGAACCAC	5' of *CUL3*	
Exon01	GACCTAAAATCATTAACATC	Exon 5 of *CUL3*	
Exon02	GCACTGCCTTGACAAATCAA	Exon 6 of *CUL3*	
Exon03	CTTACCTGGATATAGTCAAC	Exon 7 of *CUL3*	
Non-targeting 1	AACCACGGCATTGAGAGGTG	n/a	
Non-targeting 2	TACATGGTATAGTGTTTATT	n/a	
Non-targeting 3	GGGCAGAAGTTGCTGTCCTG	n/a	

Table S5: Genomic and barcode primers for targeted indel sequencing

sgRNA	Indel PCR1 forward primer (5' to 3')	Indel PCR1 reverse primer (5' to 3')
V01	ccatctcatccctgcgtgtctccTGAAGTCCAGACATTTTGTTGC	cctctctatgggcagtcggtgatgCTGTCTTGGCCCTATCCTCA
V02	ccatctcatccctgcgtgtctccAGGAAGAGAGACCAGAGTTAGCA	cctctctatgggcagtcggtgatgTGGGAGATCCAAGGTTGAAG
V03	ccatctcatccctgcgtgtctccGCTGGCACATTTTAGTGCAA	cctctctatgggcagtcggtgatgGACCCATCTCCTTTGGATGA
V04	ccatctcatccctgcgtgtctccTGCTTGTTTTATAGGCCAAGTCT	cctctctatgggcagtcggtgatgGGCTGGATGGTCCTGTCTT
V05	ccatctcatccctgcgtgtctccCATGAGTTCACCCCTTCCAG	cctctctatgggcagtcggtgatgTATCAGCAGCGTGAAAATGG
V06	ccatctcatccctgcgtgtctccCCCCCAATTCAATTATCTCC	cctctctatgggcagtcggtgatgTGGAGTGGAGCTGAGTCTTG
V07	ccatctcatccctgcgtgtctccTAGTGCACCACAGCTTCACC	cctctctatgggcagtcggtgatgCAAAGTTGGCAGCTGGTTATATT
V08	ccatctcatccctgcgtgtctccGAAATAACTCAGAACAAAACCTAATCA	cctctctatgggcagtcggtgatgGCCTTATGACCAGGAACCTTT
V09	ccatctcatccctgcgtgtctccTCTGTCCGATTGCTAGTTCG	cctctctatgggcagtcggtgatgTGGGTGTCAAATCTGGTTCA
V10	ccatctcatccctgcgtgtctccGCAAGTATGCCCAGTTCGTT	cctctctatgggcagtcggtgatgTTTGGCATTACGTTGAGTCG
V11	ccatctcatccctgcgtgtctccCGGTTTGCTCTCTGTTGCTT	cctctctatgggcagtcggtgatgGGAATGCTCCGTGGTCATAA
V12	ccatctcatccctgcgtgtctccAGCCCCTTCATCACCCTAAA	cctctctatgggcagtcggtgatgGGGTTGTAGGCCCAGTCTC
V13	ccatctcatccctgcgtgtctccCCCTAAAAGCTAGGCTGGGTA	cctctctatgggcagtcggtgatgGGGTTGTAGGCCCAGTCTC
V14	ccatctcatccctgcgtgtctccACTCTGGCGACTCCGATG	cctctctatgggcagtcggtgatgCTGCGCAGTGAGATGTTTGT
V15	ccatctcatccctgcgtgtctccCGACGGACAAACATCTCACT	cctctctatgggcagtcggtgatgTCTCTCACTCTCCGGCTCTC
V16	ccatctcatccctgcgtgtctccAGGGTCCTGGTCACATGGT	cctctctatgggcagtcggtgatgCACGCTCCTCCTCAAGACTA
V17	ccatctcatccctgcgtgtctccCTGGGACAGCAGGAGGATAG	cctctctatgggcagtcggtgatgAACTCTTCAAGTTGCAGGCTTC
V18	ccatctcatccctgcgtgtctccCTGGGACAGCAGGAGGATAG	cctctctatgggcagtcggtgatgAACTCTTCAAGTTGCAGGCTTC
V19	ccatctcatccctgcgtgtctccCAGGAAGAGACGGAGACACA	cctctctatgggcagtcggtgatgCTGGAAAGATCTCTGAAATCAAAA
V20	ccatctcatccctgcgtgtctccCACTAAATTCTGGTGTGCGTTT	cctctctatgggcagtcggtgatgAACTGTTCTGTGTCTGCACTGTC
V21	ccatctcatccctgcgtgtctccGCGCTAGCAGGAGCTGTTT	cctctctatgggcagtcggtgatgCCGGCTCATATCTGCTTCTT
V22	ccatctcatccctgcgtgtctccTGAGCAGGAATGGACACATC	cctctctatgggcagtcggtgatgGCATCTTTGACAACAAAGTGACTC
V23	ccatctcatccctgcgtgtctccGCCCTGGGGACAAGTTCT	cctctctatgggcagtcggtgatgATTTTTCCTCCCACTGCTCTG
V24	ccatctcatccctgcgtgtctccCACACAAATCTAATCTCTGGGATCT	cctctctatgggcagtcggtgatgTTCTGATTGTGGACCCTTCA
V25	ccatctcatccctgcgtgtctccTTTGTGAGACCAGCCAGAAA	cctctctatgggcagtcggtgatgTGCTCCCAAGTCCAGTCTTT
Exon01	ccatctcatccctgcgtgtctccTGGCCTTTTTAGCACTTGTCA	cctctctatgggcagtcggtgatgTCCTATTTGAGGGAGCAAGG
Exon02	ccatctcatccctgcgtgtctccTTTAC	cctctctatgggcagtcggtgatgAGAGGCGC

	ATTTTCACGGATTACCTG	AATAAGAAATGC
Exon03	ccatctcatccctgcgtgtctccTGGTTCTTCCGTTGATTTGTC	cctctctatgggcagtcggtgatgGCAGATGGAAAGCCAGAAAT

Name	Indel PCR2 primer (5' to 3')
Indel_BC_F01	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTtcagAAGTAGAGCCATCTCATCCCTGCGTGTCTCC
Indel_BC_F02	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTttcagCATGCTTACCATCTCATCCCTGCGTGTCTCC
Indel_BC_F03	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTattcagGCACATCTCCATCTCATCCCTGCGTGTCTCC
Indel_BC_F04	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTgattcagTGCTCGACCCATCTCATCCCTGCGTGTCTCC
Indel_BC_F05	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTcgattcagAGCAATTCCCATCTCATCCCTGCGTGTCTCC
Indel_BC_F06	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTtcgattcagAGTTGCTTCCATCTCATCCCTGCGTGTCTCC
Indel_BC_F07	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTatcgattcagCCAGTTAGCCATCTCATCCCTGCGTGTCTCC
Indel_BC_F08	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTgatcgattcagTTGAGCCTCCATCTCATCCCTGCGTGTCTCC
Indel_BC_F09	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTtcagACACGATCCCATCTCATCCCTGCGTGTCTCC
Indel_BC_F10	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTttcagGGTCCAGACCATCTCATCCCTGCGTGTCTCC
Indel_BC_F11	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTattcagGTATAACACCATCTCATCCCTGCGTGTCTCC
Indel_BC_F12	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTgattcagTTCGCTGACCATCTCATCCCTGCGTGTCTCC
Indel_BC_F13	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTcgattcagAACTTGACCCATCTCATCCCTGCGTGTCTCC
Indel_BC_F14	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTtcgattcagCACATCCTCCATCTCATCCCTGCGTGTCTCC
Indel_BC_F15	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTatcgattcagTCGGAATGCCATCTCATCCCTGCGTGTCTCC
Indel_BC_F16	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTgatcgattcagAACGCATTCCATCTCATCCCTGCGTGTCTCC
Indel_BC_F17	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTtcagCGCGCGGTCCATCTCATCCCTGCGTGTCTCC
Indel_BC_F18	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTttcagTCTGGCGACCATCTCATCCCTGCGTGTCTCC
Indel_BC_F19	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTattcagCATAGCGACCATCTCATCCCTGCGTGTCTCC
Indel_BC_F20	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTgattcagCAGGAGCCCCATCTCATCCCTGCGTGTCTCC
Indel_BC_F21	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTcgattcagTGTCGGATCCATCTCATCCCTGCGTGTCTCC
Indel_BC_F22	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTtcgattcagATTATGTTCCATCTCATCCCTGCGTGTCTCC
Indel_BC_F23	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTatcgattcagCCTACCATCCATCTCATCCCTGCGTGTCTCC
Indel_BC_F24	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTgatcgattcagTACTTAGCCCATCTCATCCCTGCGTGTCTCC

Indel_BC_R01	CAAGCAGAAGACGGCATACGAGATCATGATCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R02	CAAGCAGAAGACGGCATACGAGATAGGATCTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTtCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R03	CAAGCAGAAGACGGCATACGAGATGACAGTAAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R04	CAAGCAGAAGACGGCATACGAGATCCTATGCCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTgatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R05	CAAGCAGAAGACGGCATACGAGATTCGCCTTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTcgatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R06	CAAGCAGAAGACGGCATACGAGATATAGCGTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTtcgatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R07	CAAGCAGAAGACGGCATACGAGATGAAGAAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTatcgatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R08	CAAGCAGAAGACGGCATACGAGATATTCTAGGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTgatcgatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R09	CAAGCAGAAGACGGCATACGAGATCGTTACCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTcgatcgatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R10	CAAGCAGAAGACGGCATACGAGATGTCTGATGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTtcgatcgatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R11	CAAGCAGAAGACGGCATACGAGATTTACGCACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTatcgatcgatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R12	CAAGCAGAAGACGGCATACGAGATTTGAATAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R13	CAAGCAGAAGACGGCATACGAGATTCCTTGGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTtCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R14	CAAGCAGAAGACGGCATACGAGATACAGGTATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R15	CAAGCAGAAGACGGCATACGAGATAGGTAAGGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTgatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R16	CAAGCAGAAGACGGCATACGAGATAACAATGGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTcgatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R17	CAAGCAGAAGACGGCATACGAGATACTGTATCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTtcgatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R18	CAAGCAGAAGACGGCATACGAGATAGGTCGCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTatcgatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R19	CAAGCAGAAGACGGCATACGAGATAGGTTATCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTgatcgatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R20	CAAGCAGAAGACGGCATACGAGATCAACTCTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTcgatcgatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R21	CAAGCAGAAGACGGCATACGAGATCCAACATTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTtcgatcgatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R22	CAAGCAGAAGACGGCATACGAGATCTAACTCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTatcgatcgatCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R23	CAAGCAGAAGACGGCATACGAGATATTCCTCTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCTCTATGGGCAGTCGGTGATg
Indel_BC_R24	CAAGCAGAAGACGGCATACGAGATCTACCAGGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTtCCTCTCTATGGGCAGTCGGTGATg

Table S6: Chromatin immunoprecipitation-droplet digital PCR (ChIP-ddPCR) primers

sgRNA	ChIP-ddPCR forward primer (5' to 3')	ChIP-ddPCR reverse primer (5' to 3')
V01	GCACTTGGAATCCACATGAA	TCACTGTCTTGGCCCTATCC
V02	ATAGCAAACTCAGCCCCATT	GCATCTGGTGAGAGCCTTCT
V03	GCTGGCACATTTTAGTGCAA	TGGCAATCCACTCTTCTTCA
V04	GTGCACCGAATTGAAGACAG	TGGCTGTGGCTTTTATATGCT
V07	TAGTGCACCACAGCTTCACC	GCCCCTCTGAAAAGCACATA
V14	GGCTCGGCTCCCTTTATC	GAGAAGGAGGAGGAGGAGGA
V15	TCCTCCTCCTCCTCCTTCTC	TCTCTCACTCTCCGGCTCTC
V19	TGAGAGAGGGAGGAAAAAGGA	ATCTGGGCCACTCACAGAAC
V24	TCCTTGCTGATTTTGTGTTCC	CCCCTGTAGCCATCTGAGTG
V25	TGGTTAGAGCACCAGGAACC	CTTCTTGCTCCCAAGTCCAG
V10	GCAAGTATGCCCAGTTCGTT	GTCGTACCCTTGCGATGTTT
V13	GAGGCAATCCTGCACAAGAG	GGAATGCTCCGTGGTCATAA
V12	AGCCCCTTCATCACCCTAAA	CGGAGTCCTCCCTGTGTG
V13	GAAACCCCACGTGAAAAGTT	GGGTTGTAGGCCCAGTCTC
V16	AGGGTCCTGGTCACATGGT	CGCTCCTCCTCAAGACTATCC
V17	CTGGGACAGCAGGAGGATAG	CCACATGCCCTAGAAAAACA
V18	CCGAACTGGTTCCTGTGTTT	CTGCAGCTAACTCCTGCACA
NegRegion1	ATGTTGCCCAGAAACTCCTC	ATTTGACTGGGCCACAAGG
NegRegion2	AATGGAATGTGGGCAGAAGT	CAATGGGGGAGAAAATCTGA
PosRegion1	ACTAAACAGCATGCCCTTCC	CCTCTCCCCCTTCAGGATAC
PosRegion2	GCATGAGCTTCAGCTCTCTCA	TCGCAATTGAACTCCATCTC

Table S7: ChIP antibodies and optimized concentrations

Antibody	Manufacturer	Product number	Antibody/10^6 cells (uL)
p300 (EP300)	Millipore	05-257	1.2 uL
CTCF	Millipore	17-10044	2 uL
ZNF263	Abcam	ab56831	1 uL
FOS	Cell Signaling Technologies	2250S	1 uL
JUN	Cell Signaling Technologies	9165S	1 uL
YY1	Cell Signaling Technologies	2185S	2 uL
H3K4me2	Millipore	17-677	0.5 uL
H3K4me3	Millipore	04-745	0.5 uL
H3K27Ac	Millipore	17-683	0.5 uL
IgG	Millipore	12-370	1 uL